AN ABSTRACT OF THE THESIS OF

Lingxiu Zhong for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on April 7, 1992.

Title: Dependence of Secondary Structure of Biopolymers on Environment: A Circular Dichroism Study of Equivocal Amino Acid Sequences in Proteins and of Left-Handed DNA

Redacted for Privacy

Abstract Approved: _____
Dr. W. Curtis Johnson, Jr.

Studies are presented here that explore the relationship between environment and sequence specific conformation of biopolymers using circular dichroism (CD) spectroscopy. In the first part of this thesis, the way the environment affects secondary structure formed by a peptide of defined amino acid sequence is studied. Our working hypothesis is that the environment is a determinant in the secondary structure formed by a peptide. This was investigated by placing peptides of defined sequence in nonaqueous solvents as a model for the environmental effects on protein folding. We found that somatostatin, which has both α-helix and β-sheet forming potential, can form either structure depending on the solvent system. This indicates that environment can affect secondary structure of a peptide. Three equivocal

peptides are synthesized which are predicted to form α-helical structures from amino acid preference, but are found to be primarily β-strand from X-ray diffraction of their respective proteins. In some solvent systems we recover the α-helical structure predicted by amino acid preference, while in other systems we mimic the interior of the protein and produce a β-strand. These results provide experimental proof that the environment is important in determining the secondary structure formed by a peptide; therefore schemes that predict secondary structure from amino acid sequence alone can never be totally successful.

The second part of this work presents research concerning interrelationships among variants of the Z conformation in poly[d(Gm^{5}C).d(Gm^{5}C)] as a function of environment (solvent and ion content). We found that this polymer can be titrated from the Z-form in 30% ethanol to the Z'-form by adding ethanol, or stabilized as the Z'-form in 30% ethanol with divalent ions. Singular value decomposition of the CD spectra shows that the Z to Z' transition induced by ethanol or divalent ions is a two-state system. Titration of this polymer with ethanol or transition metals produces a single step titration curve; in contrast, titration with alkaline earth metals shows a complex biphasic titration curve, indicating two binding sites.

Dependence of Secondary Structure of Biopolymers on Environment: A Circular Dichroism Study of Equivocal Amino Acid Sequences in Proteins and of Left-Handed DNA

by

Lingxiu Zhong

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed April 7, 1992

Commencement June 1992

APPROVED:

Redacted for Privacy

Professor of Biochemistry and Biophysics in charge of major

Redacted for Privacy

Head of Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented April 7, 1992

Typed by Lingxiu Zhong for Lingxiu Zhong

DEDICATION

This dissertation is dedicated especially to my best friend and husband Daming Liu, for his love, support, understanding and endurance: his contribution was immeasurable; to my beautiful daughter Briana Linda Liu, who is my greatest source of joy and comfort; to my parents for their love and faith.

ACKNOWLEDGMENTS

First and most of all I would like to express my sincerest thanks to my major professor Dr. W. Curtis Johnson, for his guidance, encouragement and support. His lessons in and out of the laboratory have been greatly appreciated. The graduate experience of the past five years have been enormous and exciting, which will benefit me all my life.

My appreciation is extended to the other members of my doctoral committee, Dr. Sonia R. Anderson, Dr. Robert R. Becker, Dr. Lorraine T. Miller (graduate council representative), Dr. Henry W. Schaup, and Dr. Enoch W. Small, for their efforts in guiding my graduate program.

I would also like to thank Drs. Mark Harder, Erwann Loret, Araz Toumadje, and Mrs. Jeannine Lawrence, for their help and insightful discussions. I also thank her for proofreading this dissertation. It was a pleasure to work with them and Dr. Johnson. My thanks also go to Dr. Dean Malencik for his help in using the HPLC.

Special thanks to my dear friend and mentor, Dr. Marjorie G. McBride, for her kindness, help, encouragement and friendship. Our weekly conversations, with humor and American idioms, have enriched my family and my life in this country.

Finally the love and support of my family, my husband Daming Liu, my daughter Briana L. Liu and my parents was deeply appreciated, and is the foundation which brought this research to fruition.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This thesis uses a manuscript format. It is based on two manuscripts written for publication. The first manuscript is in press for publication in Proc. Natl. Acad. Sci. USA (1992), 89. The second manuscript has been published in Biopolymers (1990), 30, pp. 821-828. For the first manuscript in this thesis, two sections including CD studies of somatostatin and a peptide sequence from HIV Tat protein in different solvent systems were added as supporting evidence for our conclusions. The Material and Methods, and Result and Discussion sections have also been rewritten in more detail. The remainder of the sections and the second manuscript have undergone only the minimum changes necessary to conform to thesis requirements.

Dependence of Secondary Structure of Biopolymers on Environment: A Circular Dichroism Study of Equivocal Amino Acid Sequences in Proteins and of Left-Handed DNA

SECTION I

INTRODUCTION

Circular dichroism (CD) spectroscopy has been widely used to investigate the conformations of biopolymers in solution. CD measures the difference in absorption between left- and right-circularly polarized light by an asymmetric molecule. It results from the interaction of neighboring chromophores of a biopolymer and is extremely sensitive to the secondary structure of a molecule.

Biopolymers are flexible molecules that can assume a number of different secondary structures. The polypeptides of enzymatic proteins form a globular tertiary structure that contains some or all of these secondary structures: α-helix, β-sheet, β-turn and random coil. Structural proteins are often found in one of those secondary structures, usually a helical (α-helix or triple-strand helix) or β-sheet structure. DNA displays the well-known double-stranded helical structure in a number of distinct right-handed forms, designated A, B, C, and D, as well as the newly-discovered left-handed form, designated Z. The specific conformation that DNA adopts appears to depend mainly on its primary structure and its solvent conditions. RNA is confined to the closely related right-handed A-forms and the left-handed Z-form. Polysaccharides can form helical structures or have their chromophores

randomly oriented. CD is a simple and useful technique that can investigate these various forms in solution, including water, the biological solvent. It can be used to identify the various secondary structures in DNA or polypeptides, and quantitatively analyze the fractions of different secondary structures in proteins.

For proteins, different secondary structures display different CD spectra, as Figure 1.1 shows (Brahms and Brahms,1980). An α-helical conformation has a highly distinctive CD spectrum, with an intense double minimum at 222 and 208-210 nm, and a more intense maximum near 190-195 nm. A β-strand CD is low in intensity and exhibits a variety of shapes. It is characterized by the absence of the features characteristic of an α-helix: it usually has a single negative band at 215-220 nm and a positive band at 195-200 nm. It also has a second negative minimum in the 170-180 nm range, and a cross-over from positive to negative above 185 nm. CD spectra of the many types of β-turns are still not well characterized. The most common CD has a positive band around 205 nm and a negative one below 190 nm plus a weak, negative one near 225 nm. The unordered form (or "random coil") shows a strong negative CD band near 200 nm and a weak positive or negative one near 220 nm. In principle, CD is also able to distinguish among protein tertiary structure classes: all-α (predominantly α-helical), all-β (predominantly β-sheet), α+β (separate α-helix- and β-sheet-rich regions), and α/β (intermixed segments of α-helix and β-sheet) (Manavalan and Johnson, 1983).

Different forms of DNA also display different CD spectra, as shown in

Figure 1.2 (Riazance et al., 1985). B-form DNA has a positive band around 280 nm, a crossover around 260 nm, and a strong positive band near 190 nm. The A-form has a more intensive band at 270 nm, a negative band around 210 nm, and also a strong positive band at 190 nm. In contrast, the left handed Z-form shows an almost mirror image CD of right-handed DNA with a negative band around 295 nm and a deep negative valley at 194 nm and a positive band below 185 nm. All of these biopolymers can be denatured so that their secondary structure is lost, and the chromophores assume a random orientation limited only by van der Waal's interactions. One secondary structure can be induced into another secondary structure by changes in environment. We can also use CD to follow the dynamic conformational changes of these flexible molecules. With this technique we can study the relationship between environment and conformation with specific sequences of these biopolymers, such as changes in secondary structure due to ligand binding, solvent variation, temperature or pH change, etc.

Here I report two major research projects involving CD to study the environmental effects on the conformation of proteins and DNA. One project concerns protein folding, another concerns left-handed DNA. Both of these topics have been the focus of intensive research recently. In the first project we studied how the environment affects secondary structure formed by oligopeptides, while the second project involved the Z'-form of poly[d(Gm^{5}C).d(Gm^{5}C)] in different environments. In the next two sections some background material that is relevant to these two projects will be reviewed, and the significance of the research will be discussed.

Section IA starts with a literature review of the protein folding problem approached by predicting secondary structure from primary structure, followed by a discussion of the difficulty with these methods. Then the rationale for the research involving solvent systems and protein folding will be reviewed. Finally the significance of studying environment effects on secondary structure in proteins will be presented.

An overview of DNA conformation will be summarized in Section IB. Different left-handed DNA structures in the Z' family will be compared, then structural studies on methylated DNA with a Gm^5C sequence will be discussed. Finally, specific aims of studying the Z'-form of poly[d(Gm^5C).d(Gm^5C)] will be given.

Figure 1.1 The CD for various secondary structures: α-helix (——), β-strand (–⌣–⌣), β-turn (····), and random coil (- - -), redrawn from Brahms and Brahms (1980).

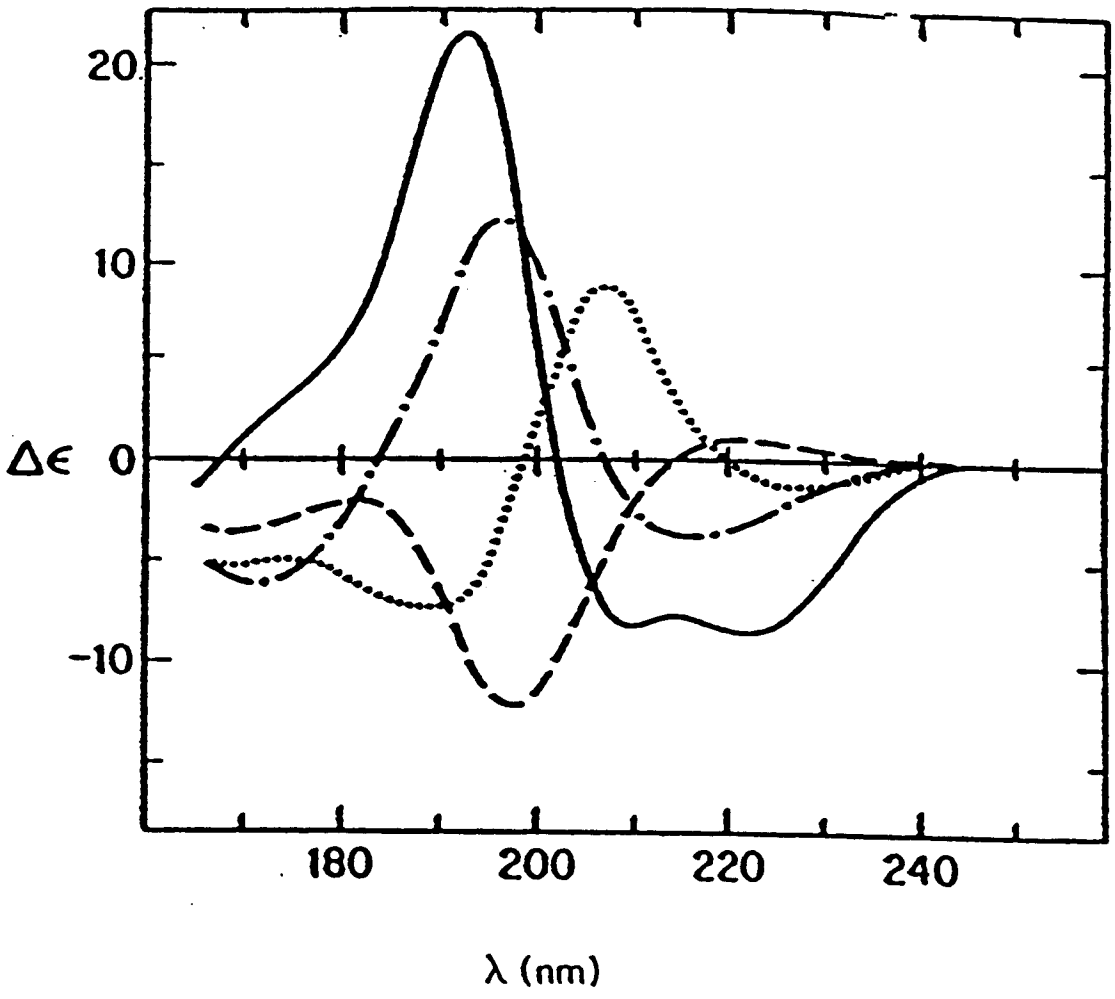

Figure 1.1

Figure 1.2 The CD for various secondary structures of poly[d(GC).d(G-C)]. At 22 °C as the B-form in 10 mM sodium phosphate buffer, pH 7 (...), as the A-form in 80% 2,2,2-trifluoroethanol (TFE), 0.67 mM sodium phosphate buffer, pH 7 (---), and as the Z-form in 2 M sodium perchlorate, 10 mM sodium phosphate buffer, pH 7 (——), redrawn from Riazance et al. (1985).

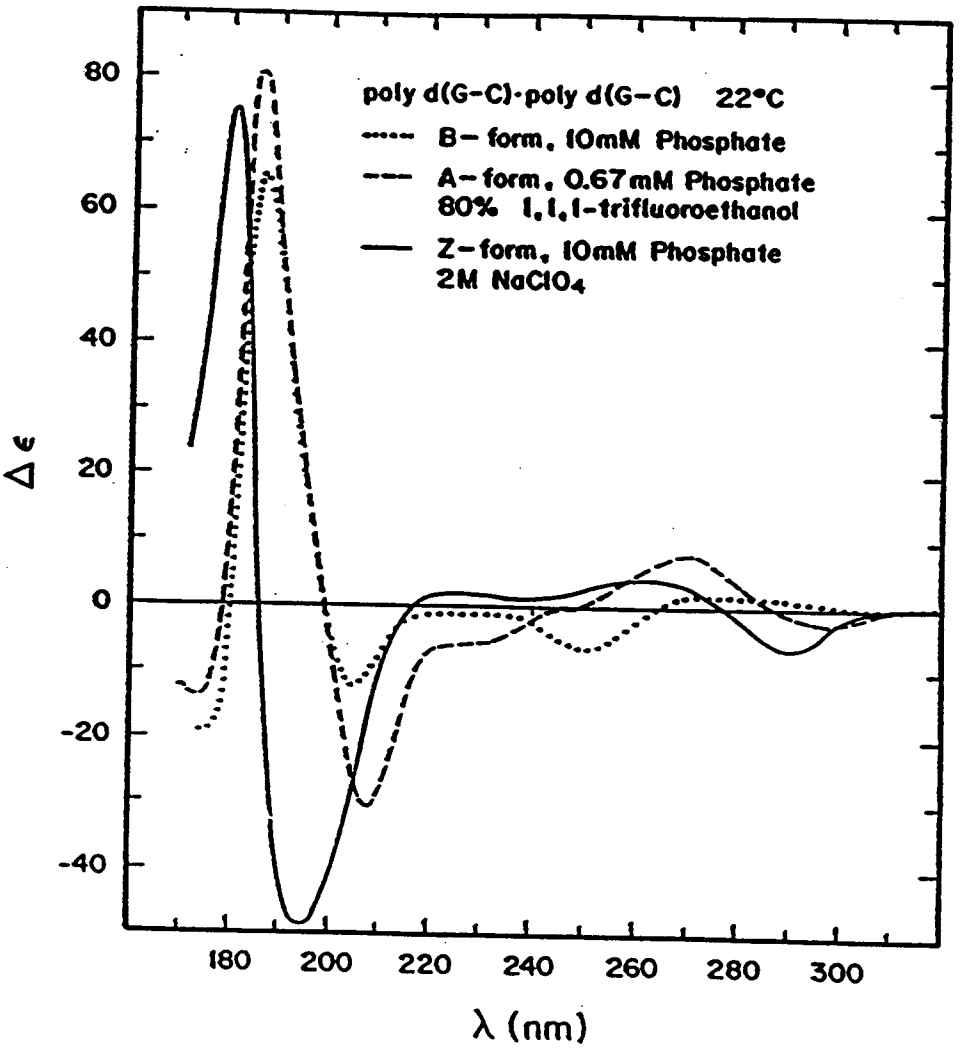
poly d(G-C)-poly d(G-C) 22°C
B- form, 10mM Phosphate
A- form, 0.67mM Phosphate
80% 1,1,1-trifluoroethanol
Z- form, 10mM Phosphate
2M NaClO4
80
60
40
20
0
-20
-40
Δε
180
200
220
240
260
280
300
λ (nm)

Figure 1.2

A. Dependence of Secondary Structure of Proteins on Environment

1. History of Prediction Methods for Secondary Structures

The process of protein folding has been studied for many years. Proteins were known as chains of covalently linked amino acids by Emil Fischer and Franz Hofmeister in 1902 (Haschemeyer and Haschemeyer, 1973). Later the process of denaturation was discovered and distinguished from the process of aggregation (Chick and Martin, 1911), and was considered to be either hydrolysis of the peptide bond (Wu and Wu, 1925; Anson and Mirsky, 1925) or dehydration of the protein (Robertson, 1918). Wu first pointed out that protein denaturation was an unfolding process (1929, 1931). Then Mirsky and Pauling (1936) suggested a structural theory of proteins to explain their ability to be denatured. Specific proposals for structural elements emerged 15 years later with models for the α-helix, and parallel and antiparallel pleated sheets (Pauling et al., 1951; Pauling and Corey, 1951). These predictions were confirmed by the first successes in determining protein structure through X-ray crystallography (Kendrew et al., 1960; Blake et al., 1965; Perutz et al., 1965; Kartha et al., 1967; Matthews et al., 1967). At the same time, Anfinsen et al. (1961) showed that under the right conditions a denatured protein could be refolded to regain its enzymatic activity. This was an experimental demonstration that the amino acid sequence in a protein contained enough information to determine its native structure, and signaled the beginning of the protein folding problem.

The distinguishing feature of proteins is their ability to adopt a unique

tertiary structure made up of secondary structures. It is this precise three-dimensional structure which allows a protein to play particular functional roles. A protein folded in one specific way may create a binding site for another molecule and a series of reactions will occur. Otherwise, the binding site and reactions will not exist. But what exactly are the physical forces involved in this folding? How can we predict whether a certain protein will fold in one specific way rather than so many others?

Empirical approaches to the protein folding problem generally look to protein structures as determined by X-ray crystallography for information about how the amino acid sequence of a protein determines its ultimate structure. In the hierarchic model, the primary sequence folds into repeating secondary structures, which in turn coalesce into a tertiary structure. The first step in such a model is to establish how the sequence determines secondary structure. As Richardson and Richardson (1988) have stated, "The empirical preferences of each amino acid for the different types of secondary structure have formed a central theme of the efforts to predict three-dimensional protein structure from amino acid sequence."

A description of a protein's three-dimensional structure has been proposed to include the classification of globular proteins based on the predominant type of secondary structure. Levitt and Chothia (1976) classified the tertiary structures of proteins into five types: all-α, all-β, α/β, $\alpha+\beta$, and irregular. With the relationship among certain tertiary structures established, Richardson (1981) presented a classic treatise that simplified structure by demonstrating the relationships among the proteins for which X-ray structural

data were available. This has simplified the task of empirically determining tertiary structure by packing the secondary structures (Ptitsyn and Finkelstein, 1979; Cohen et al., 1979, 1982, 1983; Chothia et al., 1977; Efimov 1979; Rose 1979; Lifson and Sander, 1980). Thus the prediction of secondary structure is an integral step in predicting the overall three-dimensional structure of a protein. Protein secondary structure refers to the local regular arrangements of the polypeptide chain backbone which are stabilized by hydrogen bonds between peptide amide and carbonyl groups (Schulz and Schirmer 1979). In general, secondary structure in proteins has been classified into four major types: α-helix, β-sheet, β-turn, and random coil. A more detail description of various secondary structure can been found in the review by Richardson (1981).

Prediction of secondary structure from primary structure began with the α-helix. Researchers found that groups of amino acids were helix formers or disrupters (Blout, 1962; Guzzo, 1965), and there was a correlation between amino acid composition and protein structure (Davies, 1964). Prothero (1966, 1968), Cook (1967), Periti et al. (1967), and Low et al. (1968), all examined the X-ray structures of proteins and tried to establish the recognition of α-helices from their amino acid sequences. The helical wheel, devised by Schiffer and Edmundson (1967) to show the contribution of hydrophobic side chains to the stability of an amphophilic helix, is still in use today. Ptitsyn (1969) made a statistical analysis of the distribution of different amino acid residues among helical and nonhelical regions of seven globular proteins. It was discovered that different kinds of amino acids have a different tendency to

occur in different regions of α-helix or other structure.

More recent schemes that make use of amino acid preferences and stereochemical criteria taken from crystal structure data have been fairly successful. Over twenty methods of predicting secondary structure from amino acid sequence have been proposed. These predictions mostly assume that the local sequence (short-range interactions) determines local structure.

In general, we can divide these methods into two categories. First are the statistical methods. These methods usually assign one type of secondary structure: α-helix, β-strand, random coil or β-turn to each amino acid in a protein with known structure. The parameters for the probability of an amino acid being in each type of secondary structure are derived based on the distribution of amino acids in these types. The methods vary both in how the parameters are derived and how they are then used to predict structure. One of the most popular prediction methods is proposed by Chou and Fasman (1974a,b; 1978a,b). They conducted a statistical survey of 15 proteins and established conformational potential for α-helix, β-sheet and β-turn for all 20 amino acids; empirical rules were then derived to determine the secondary structural regions in the proteins. In attempts to improve the Chou and Fasman method, Garnier et al. (1978) published a method which considered the influence of chain length-composition on protein conformation.

Second, a few methods are based on stereochemical criteria. Lim has taken into account both the hydrophobicity and size of side chains to propose favorable patterns of residues that will form α-helices and β-strands. This method emphasizes the importance of the positions of hydrophobic and

hydrophilic residues and also incorporates, to some extent, long-range interactions. The most frequently used methods to date have been the empirical approaches of Chou and Fasman (1974a,b; 1978a,b) and of Robson and coworkers (Garnier et al., 1978), and the stereochemical method of Lim (1974a,b).

Some researchers combined several predictions to obtain joint predictions for their proteins (Schultz et al., 1974; Matthews, 1975; Manavalan et al., 1986). The joint predictions were shown to be comparable with the best single prediction. Recently, Nishikawa and Noguchi (1991) showed that their joint prediction method is an improvement over individual methods by 2 to 5%.

With the large crystal structure database available, there are now predictions of secondary structure based on homology, see for instance Pongor and Szaley (1985), Sweet (1986), Nishikawa and Ooi (1986), Levin et al. (1986), and Zvelebil (1987).

The limits of prediction accuracy were discussed by Palau et al. (1982), who recognized that it might be advantageous to consider the various types of tertiary structure separately. Both linear optimization of predictors for secondary structure (Edelman and White, 1989) and neural networks (Qian and Sejnowsky, 1988; Holley and Karplus, 1989) have been used to reach the limits of predicting secondary structure from the local information in the primary structure.

Although numerous prediction methods have been developed, their best accuracy has been only around 60% (Nishikawa and Noguchi,1991). The most popular methods for predicting protein secondary structure give a

prediction accuracy of 50-56% from independent tests (Popoz, 1980; Nishikawa, 1983; Kabash and Sander, 1983; Yada et al., 1988). This will be discussed in the next section.

Site-directed mutagenesis offers the opportunity to see how changing an amino acid changes the structure of a protein. Clearly, this is another empirical way of unraveling how the primary structure of a protein determines its secondary and tertiary structure. A number of researchers have taken up this idea, among them Shortle (1989), Matthews (Wozniak et al., 1990), and Schellman (1981, 1987). Mutations need to be judiciously chosen, for the same reason that other approaches find the problem complicated; the number of possibilities for making a protein of 100 amino acids with 20 possibilities at each point exceeds the number of atoms in the universe. Perhaps the most important result of this effort is the realization that proteins are extremely tolerant of mutations, since most substitutions turn out to be allowed without drastic changes in structure (also see Bowie et al., 1990). About two decades ago a protein was considered to be a rigid structure with closely packed side chains, and it was difficult to understand how an amino acid substitution could be tolerated in the interior of a protein. Today it is known that the interior of a protein is flexible, and this could also indicate that an α-helix can change to a β-sheet. This suggests that long-range sequence is important, leading us to our research.

2. "Second Genetic Code" Remains to be Solved

The first half of the genetic code, deciding how a sequence of DNA

bases is translated into a sequence of peptides, was depicted more than 20 years ago. The second half of the genetic code, determining how a linear amino acid sequence folds into a protein, has been unsolved for that long a time. The ultimate solution to the "protein folding problem" will be the elucidation of the "second genetic code" relating the amino acid sequence of a protein to its secondary, tertiary, and quaternary structures. Numerous attempts to solve this problem have been published (for review see Fasman, 1989). Many laboratories now believe that the problem is tractable, and recently it has been the subject of intense research. It is much easier to sequence a protein than to determine its structure through X-ray crystallography or NMR techniques. With so many sequences known, it would be nice to predict the structure of these proteins. Furthermore, protein engineering and de novo design require an understanding of how the product will fold, making this a central problem in molecular biology.

Our interest in protein folding stems from using CD spectra of proteins to predict secondary structure (Hennessey and Johnson, 1981; Manavalan and Johnson, 1987). When we predicted the fractions of different secondary structures for Eco. RI endonuclease (Manavalan et al., 1984), and thymidylate synthase (Manavalan et al., 1986), we also predicted the locations of the secondary structures by using a combination of predictions from primary structure. Subsequently, the crystal structures for these two proteins appeared (McClarin et al., 1986; Hardy et al., 1987) and the CD predictions were quite good. However, one sequence in Eco. RI endonuclease (residues 103 to 115), which was predicted to be α-helix from primary structure, turned

out to be β-strand. Another sequence in thymidilate synthase (residues 53 to 68), which was predicted as β-strand turned out to be α-helix. This is not unusual, for as Cohen (1983) says, "... attempts so far have invariably misassigned one or more of the secondary structure elements in all the proteins studied."

Kabsch and Sander (1983) examined various methods for the prediction of secondary structure from primary structure, but found for 62 proteins that none of the methods predicted better than 56% of the residues correctly as α-helix, β-strand and β-turn. Nishikawa (1983) also showed error levels of more than 40% in several representative prediction methods. These large errors certainly arise from the probability nature of these methods. In addition, secondary structure formation in proteins is not totally dependent on the sequence information alone, long-range interactions also play certain roles (Sternberg and Thornton, 1978; Schulz and Schirmer, 1979; Nishikawa and Noguchi, 1991).

Furthermore, we have carried out CD studies that demonstrate changes in secondary structure with the binding of ligands (Manavalan et al., 1985; Manavalan et al., 1986), changes in secondary structure with solvent variation (Hennessy et al., 1987), and changes in secondary structure with quaternary interaction (Hennessey et al., 1982). The change in secondary structure with quaternary interaction is particularly startling, with 8% β-sheet and 7% other (random) structure being converted to α-helix. One aspect of such interactions is to change the effective solvent that a peptide sequence sees, then the secondary structure preference of a peptide sequence may well

depend upon its environment.

3. Solvent Systems and Protein Folding

It has long been recognized that solvent is important in peptide folding. For instance, Fasman in Poly-α-amino Acids (1967) can be quoted, ". . . interaction of solute-solvent, always plays an important role in determining structure." Jirgensons (1977, 1981) showed that sodium dodecylsulfate (SDS) is effective in disorganizing tertiary structure and enhancing the α-helical content of many proteins. He called this reconstructive denaturation. Using various organic solvents, Tanura and Jirgensons (1980) were able to reorganize a variety of proteins and increase either their α-helical or β-strand structure. The structure of gramicidin also depends on its environment (Short et al., 1987; Killian et al., 1988). Kubuta et al. (1983a,b) have used SDS to form both α-helical or β-strand structure in certain alternating copolymers. Many bioactive peptides without disulfide bonds or prosthetic groups exist in aqueous solution as a random structure (Burgen et al., 1975). However, in hydrophobic solvents and those less polar than water, some of these peptides can assume an ordered conformation. Prediction of secondary structure from primary structure indicates that glucagon has the potential to form both α-helical and β-structure (Chou and Fasman, 1975), with α-helix forming in organic solvents (Gratzer and Beaven, 1969; Epand, 1972; Contaxis and Epand, 1974), and lipids (Shneider and Edelhoch, 1972; Epand et al., 1977), and β-structure as aggregated long fibrils developing in acid solution (Gratzer et al., 1968). Gierash and coworkers have shown that signal peptides

involved in protein export are α-helical when inserted in lipid monolayers, but have a β-structure on the surface (Briggs et al., 1986; Cornell et al., 1989). Angiotensin II, which is a random coil in aqueous solution, can form β-sheet in solvents such as 2,2,2-trifluoroethanol (TFE) (Devynck, et al., 1973). Likewise, endorphin becomes helical in methanol (Yang, et al., 1977), TFE (Graf, et al., 1977), and SDS solution (Yang, et al., 1977). Clearly, organic solvents, lipids, or surfactants provide a hydrophobic environment which enhances the intramolecular hydrogen bonding and nonpolar interactions between amino acid residues; this in turn stabilizes the ordered structure of polypeptides. In globular proteins, stabilizing factors are provided by hydrophobic interaction, hydrogen bonds, salt bridges, and the long-range interaction of side chain groups.

4. Specific Aims of the Research

The conventional view is to look at amino acid preferences for the different secondary structures, but our research took a different tack. We believe that the unusual sequences that do not form the expected secondary structure are the interesting ones. Perhaps the environment that a secondary structure sees after a protein folds into its tertiary structure causes the secondary structure to change. This idea is implicit in the notion that the secondary structure preference for amino acids in membrane proteins is different from the amino acid preferences in soluble proteins (Wallace et al., 1986). This can be investigated by looking at peptide sequences in nonaqueous solvents as a model for the environmental effects on protein

folding.

Certainly the study of polynucleotides in nonaqueous solution has proved particularly valuable. DNA is polymorphic, and the various secondary structures that it achieves can all be found by varying the properties of the solvent and the sequence of the polydeoxynucleotide. The normal B-form of natural DNA is found in aqueous solution at moderate salt (Brahms and Mommaerts, 1964; Tunis-Schneider and Maestre, 1970). Increasing salt or adding alcohols decreases the number of base pairs per turn to mimic the DNA that is wound around histone cores (Baase and Johnson, 1979). Solvents that are primarily ethanol produce A-form DNA (Ivanov et al., 1973; Girod et al., 1973). Mixtures of ethanol with a low water content produce a bizarre form of denatured DNA termed the P-form (Zehfus and Johnson, 1981). High concentrations of ethanol can produce the C-form of DNA when lithium salt is present (Bokma et al., 1987). Ethanol and high salt produce the left-handed Z-form for certain deoxypolynucleotides with alternating purine-pyrimidine sequence (Pohl and Jovin, 1972). We anticipate that studies of peptide sequences in nonaqueous solution will produce the same type of useful information.

To check our working hypothesis that the environment determines the secondary structure formed by an amino acid sequence, we started with the investigation of a bioactive peptide, somatostatin, in different solvent systems. We found that it could form the predicted secondary structure in different solvent systems. This shows that environment can affect the secondary structure of an amino acid sequence. If environment from the tertiary

structure causes anomalous secondary structure, then it should be possible to recover the expected secondary structure by removing the "anomalous" sequence from the protein. Our research results prove that this is true. We sliced out three such equivocal amino acid sequences that are predicted to be α-helix but found as β-strand in well-defined proteins. We were able to recover the predicted α-helix in some solvent systems. Taking the idea one step further, we found solvent systems that mimic the interior of the protein so that these equivocal amino acid sequences form the anomalous secondary structure. Here we used solvent to encourage the transition between α-helix and β-form, rather than looking at transitions from random to α, or random to β. We found that the solvent plays a dominant role, relative to the intrinsic helix preferences. Finally, we studied a random structure sequence in the same solvent systems as a control, and found it to remain an irregular structure. Our research is the first experimental proof that the secondary structure of any given segment of a polypeptide chain within a protein assumes a conformation determined by the environment of the segment as well as by the amino acid preferences.

B. Z'-form of Poly[d(Gm^{5}C).d(Gm^{5}C)]

1. The Polymorphism of DNA

By 1952 DNA was recognized as the genetic substance, but the question of how it functioned remained unanswered until Watson and Crick (1953) discovered the structure of DNA. Their discovery of the double helical structure of the DNA molecule was a breakthrough for many branches of biology, because it explained how the structure of the molecule allows it to function as a template for copying genetic information.

They depicted the right-handed double helix with antiparallel strands stabilized by hydrogen bonding between bases on opposite strands. The two strands with 10 base pairs per turn and 3.4 Å per base pair are complementary, A always pairing with T and G always with C. In the Watson-Crick model the base pairs are stacked on one another with their planes nearly perpendicular to the helix axis. This was consistent with Wilkins' X-ray patterns for fibers of DNA at high humidity, the B-form (Langridge et al., 1960). The double helix with anti-glycosidic bonds has two distinct grooves: a minor groove that lies between the 1-carbon of the sugars on two strands, and a major groove on the opposite side of the helix. The more recently refined model obtained from single crystal X-rays (Dickerson, 1990) shows that it has a C_2'-endo sugar pucker. The base pairs in the fiber are about 6° off from perpendicular to the helix axis. This B-form DNA is generally considered to be the conformation of DNA under physiological conditions and expected to exist in the highly aqueous milieu of the cell.

It is now well established that DNA structure is polymorphic (Wells, 1988; Kennard and Hunter, 1989, Zimmerman, 1982; Cold Spring Harbor Symp. Quant. Biol. 1983; Saenger, 1984; Leslie et al., 1980), with the sequence-specific conformation sensitive to different environments: cation type, temperature, and solvent or humidity. This polymorphism is believed to participate in the regulation of gene expression.

The A-form of DNA occurs in fibers of the sodium salt at 75% humidity (Fuller et al., 1965, Chandrasekaran et al., 1989). It has 11 residues per turn and a rise of 2.6 Å per base pair, the diameter being a few angstroms larger than that of the B-form. The base pairs are again planar, but in this form they are tilted 20° with respect to the helix axis. The deoxyriboses are in the C'_3 endo conformation. The two grooves of the A-form helix are more nearly equal in depth. In solution, the B-form can be converted to the A-form by reducing the water activity. RNA-DNA hybrids assumes an A-form helix rather than the B-form helix even in normal aqueous buffer.

C-DNA has been observed with fiber X-ray diffraction, existing in low humidity with Li^+ salt. This duplex has 9.3 base pairs per turn and 3.3 Å per base pair (Marvin et al., 1961; Rhodes et al., 1982), a slight 6° tilt away from perpendicular to the helix axis, and a slight twist to the bases. The CD of C-form in solution has been published by Bokma et al. (1987)

There is also D-form DNA found in several nonguanine-containing deoxypolymers under relatively dehydrated conditions in fibers (Davies and Baldwin, 1963; Mitsui et al., 1970; Arnott et al., 1974). It is a conformation with 8 base pairs per turn and 3 Å per base pair. In a mixture of solvents

such as methanol, ethanol and water, DNA forms a combination of collapsed tertiary and denatured secondary structures, named the P-form (Zehfus and Johnson, 1981, 1984). CD and ^{31}P NMR also show that a double strand polynucleotide with alternating A-T or A-U bases and their modified bases can adopt an X-form in high concentrations of CsF, or ethanol with CsCl (Kypr and Vorlickova, 1988); X-DNA is a right-handed double helix with a zig-zag backbone. However, models of these conformations are not well defined as yet.

All the conformations of DNA mentioned above are right-handed. The most investigated and surprising conformation of DNA perhaps is the left-handed Z form DNA, which will be discussed next.

2. The Z DNA Family

In 1979, Alexander Rich and his colleagues carried out single-crystal X-ray diffraction studies of a hexamer with the alternating sequence d(CGCG-CG) (Wang et al., 1979). They found unambiguous proof for the existence of a left-handed double-helical DNA. Shortly thereafter, the crystal structure of the tetramer d(CGCG) was published (Drew et al., 1980). In crystal structures of both the d(CG) tetramer and hexamer, the self-complementary oligonucleotides are arranged into antiparallel, left-handed helical duplexes, such that 12 Watson-Crick-type GC base-pairs would complete a turn about every 45 Å. In this left handed DNA, the guanine is in a syn-conformation, the cytosine in an anti-conformation. The sugar puckers are respectively C'_3-endo and C'_2-endo. These alternating conformations produce an irregular zig-zag

course for the sugar-phosphate backbone, hence the Z DNA designation for this structure. The Z-form has a single, very deep helical "groove" that corresponds in location to the minor groove. There are two different environments for the phosphate groups in the G_pC and C_pG sequences. Small but significant variations in the conformations of the different crystals of alternating GC oligonucleotides suggest that there exists a family of left-handed structures (Crawford et al., 1981; Drew and Dickerson 1981; Wang et al., 1981; Drew et al., 1980). The principal variation occurs in the orientation of the phosphate group in the G_pC sequence: the phosphates are found either facing the helical groove (Z_I) or rotated away from it (Z_{II}). The latter conformation is often found when hydrated magnesium ions are complexed to a phosphate oxygen atom. This also gives rise to differences in sugar and sugar-phosphate torsional angles. These two phosphate orientations are stabilized by water-mediated intranucleotide hydrogen bonds linking the guanine N_2 amino group with oxygen at the deoxyguanosine 3'-phosphate in each $d(G_pC)$. In Z_I, one such water molecule suffices to provide this interaction while in Z_{II}, two water molecules are needed. More recently, studies have also shown binding of cobalt (III) hexamine (Brennam et al., 1986; Gessner et al., 1985), and ruthenium (III) (Ho et al., 1987) to bases and phosphate oxygens, causing Z_I to Z_{II} conversion at these sites. A related change in the conformation of the phosphate group occurs in the crystalline tetramer at high and low salt as solved by Drew et al. (1980) and Crawford et al. (1980). In this structure, labeled Z', the phosphate group of the GC sequence again rotates outward, as in going from Z_I to Z_{II}. In the Z'-DNA

double helix, the deoxyguanosine sugar is puckered C'_1-exo, a variant of C'_2-endo and contrasting with the C'_3-endo found in Z-, Z_I-, and Z_{II}-DNA. This difference in sugar puckering for Z- and Z'-DNA can also be explained by variation in solvent interactions around the guanine N_2 amino group. A water molecule bridges the N_2 amino and 3'-phosphate groups of deoxyguanose intramolecularly via hydrogen bonding. In the high salt d(CGCG), however, the water is replaced by a chloride ion in hydrogen-bonding contact with this amino group, but the Cl^{-1} repels the phosphate. This brings a change in the sugar puckering mode from C'_3-endo (Z-DNA) to C'_1-exo (Z'-DNA). By comparing the Z-DNA crystal structures with that of B DNA, Drew and Dickerson (1981) observed that base inclination and stacking are relatively insensitive to the conformation of the backbone of Z DNA, and that changes in backbone structure need not be reflected in changes in base pair interactions to the degree that would be observed in B DNA. Consequently, the energy of conversion among the left handed forms is expected to be small, and they proposed that polymorphism in the Z forms of DNA is not limited to these few variants, but that a continuum of left handed forms of DNA should exist which involves differences in backbone structure.

The discussion by Wang et al. (1981) contains the most detailed polymeric models based on the oligomers. The predicted diffraction pattern for their model of the predominant conformation (Z_I) fits the observed pattern (Arnott et al., 1980) of fiber poly[d(GC).d(GC)], which provides strong evidence for the Z conformation in fibers of this polymer.

Some years ago, Pohl and Jovin (1972) observed that the circular

dichroism (CD) spectrum of the alternating copolymer, poly[d(GC).d(GC)], underwent a novel inversion when the polymer was exposed to high salt. This is a consequence of a B to Z transition in solution. The left-handed conformation has also been observed in solution by NMR (Feigon et al., 1984) and Raman spectroscopy (Rich et al., 1984). A second solution form of poly[d(GC).d(GC)] was described by Pohl (1976), and later by Hall and Maestre (1984), who also called this the Z'-form. They discovered that it is produced at high ethanol concentration (85% v/v), and the Z to Z' transition is a function of alcohol percentage. Recently Harder and Johnson (1990) found that both multivalent ions and 85% ethanol are required to produce the original Z'-form of poly[d(GC).d(GC)] in solution. Differences among the Z' spectra produced by the different ions suggest that a least three families of Z' structure exit. The authors showed that the Z'-form of left-handed poly[d(GC).d(GC)] in solution depends on the presence of multivalents and thus can be related to the Z_{II}-form in crystals.

3. Cytosine Methylation in a GC Sequence Stabilizes Z-DNA

In eukaryotic DNA, the sequence d(Gm^{5}C) occurs quite frequently; in many organisms it composes more than half of all "d(CG)" sequences and is the major result of DNA methylation (Razin and Riggs, 1980). Methylation of CG residues is thought to be associated with gene inactivation, while subsequent removal of the methyl group is associated with gene activation (Razin and Riggs, 1980; Doerfler 1983).

Normally, the Z conformation is adopted only by regions of

polynucleotides having alternating purine-pyrimidine sequences and then only at salt concentrations much higher than that found in vivo. However, Behe and Felsenfeld (1981) showed that the Z-form can be adopted at physiological ionic strength if a substantial fraction of the C residues are methylated to form 5-methylcytosine. In their experiment, the midpoint of the transition from B to Z DNA in poly[(dGm^{5}C)d(Gm^{5}C)] required 0.6 mM Mg^{2+} in solution with 50 mM NaCl, which is a lowering of the magnesium requirement by three orders of magnitude. The presence of cations like Mg^{2+}, Na^+, or small molecules like spermine and spermidine, which are found in most cells, favors the Z-form. Much less of these molecules is required to stabilize Z-DNA in the methylated polymer. That the Z-form of the methylated polymer poly[(dGm^{5}C).(dGm^{5}C)] can be stable under physiological conditions has also been confirmed by NMR (Patel et al., 1981), IR studies (Taillandier, 1985), and X-ray fiber diffraction (Behe et al., 1980; Zimmerman, 1982).

The Z-form structure of the methylated molecule was revealed when the hexamer d(m^{5}CG)$_3$ was solved by Fujii et al. (1982) at 1.3 Å resolution. The overall structure of the molecule is quite similar to that of the unmethylated Z-DNA in the d(CG)$_3$ crystal (Wang et al., 1979). For Z-form d(m^{5}CG)$_3$, the methyl group forms a small hydrophobic patch on the surface of the molecule and causes slight changes in the twist angle between base pairs. In contrast, the position of the methyl group in B-form d(m^{5}CG)$_3$ projects into the major groove, and is surrounded by water molecules. Methylation of C causes stabilization of the Z-form relative to B-form because of the difference in environment of the cytosine methyl group in these two conformations (Fujii

et al., 1982).

Since the discovery of Z-DNA, several attempts have been made to explore the biological relevance of this DNA form in vivo (for review see Rich et al., 1984). At the structural level, the existence in eukaryote organisms of Z-DNA is expected to introduce topological changes and constraints in those regions of chromatin associated with it. In order to answer the question of how the B to Z DNA transition may affect the nucleosome, most studies took advantage of the fact the poly[(dGm^{5}C).(dGm^{5}C)] can be induced to undergo the B to Z transition under very mild conditions, which can be easily achieved under in vivo situations. Some studies using poly[(dGm^{5}C).(dGm^{5}C)] and histones were contradictory. Under some conditions nucleosome reconstitute in a B-DNA form (Nickol et al., 1982), while in other experiments Z-DNA is found in the nucleosomes (Miller et al., 1983; Prevelige and Fasman (1983). Ausio et al. (1987) reexaminated the reported B to Z DNA transition in nucleosomes reconstituted with poly[dGm^{5}C).d(Gm^{5}C)]. They found that only free nucleotide released by core particle dissociation undergoes the B to Z transition, and no evidence has been found for "Z nucleosomes" in solution. They proposed a model suggesting that all of the DNA that remains bound to the histone octamer retains the B-form.

4. Specific Aims of the Research

As mentioned above, methylation of dC residue sequences provides considerable further stabilization of left handed structures (Behe and Felsenfeld, 1981), perhaps by increasing hydrophobic interactions (Fujii et al.,

1982). It has also been associated with inhibition of transcription (Razin and Riggs, 1980). Therefore, considerable research has been carried out on the B to Z transition of poly[d(Gm^{5}C).d(Gm^{5}C)]. However, none of the research has studied the Z'-form of this polymer. Furthermore, the polymer is extremely sensitive to its environment. Trace amounts of multivalent metal ions, a slightly higher concentration of salt, increasing temperature, changes in pH, and ligand or drug binding can all promote the B to Z transition (Chen, 1986; Chaires, 1985; Devarajan and Shafer, 1986). It is very interesting to see under what conditions it will form the Z'-form. Here we investigate it with organic solvents and multivalent ions, and compare it with the Z'-form of poly[d(GC).d(GC)] so that we can shed some light on the Z'-form and the role which the methyl group plays in Z'-DNA.

Interesting results came out of our research, which showed that this polymer can form the Z'-form in 85% ethanol, or 30% ethanol with divalent ions. First, this polymer can be titrated from the Z-form in 30% ethanol to the Z'-form by adding ethanol up to 90%. The near UV CD bands, which are sensitive to the conformations of the sugar and phosphate, change markedly, while the vacuum-UV CD, which is sensitive to base-base interactions, remains almost unchanged. The titration CD spectra show a single isosbestic point, and singular value decomposition reveals there are two states in the Z to Z' transition. Second, this polymer can also be titrated to the Z'-form in 30% ethanol with divalent metals. However, the CD spectra of transition metal and alkaline-earth metal titration are quite different. The CD spectra of transition metal titration are similar to the ethanol titration, the SVD analysis

indicating only a single binding site. However, we observed heterogeneity of the active calcium and magnesium binding sites of poly[(dGm^{5}C).(dGm^{5}C)]. Titration of this polymer with calcium or magnesium in sufficient ethanol to fully stabilize the Z'-form (30% v/v) is clearly biphasic with respect to metals concentrations, although an SVD analysis of the spectra indicates only two component spectra in the mixture. The simplest explanation for this phenomenon is that there are two classes of binding sites differing in their affinity for each of these metals; but the spectra, and therefore the DNA conformations, at the two sites are indistinguishable.

SECTION II

ENVIRONMENT AFFECTS AMINO ACID PREFERENCE FOR SECONDARY STRUCTURE

Lingxiu Zhong and W. Curtis Johnson, Jr.

Department of Biochemistry and Biophysics
Oregon State University
Corvallis, OR 97331-6503

In press, Proc. Natl. Acad. Sci. USA (1992), 89.

Expanded here as explained in the Preface

ABSTRACT

Three equivocal amino acid sequences were synthesized which are predicted to be α-helical from amino acid preference, but are found to be primarily β-strand from X-ray diffraction of their respective proteins. In some solvent systems we recover the α-helical structure predicted by amino acid preference, while in other systems we mimic the interior of the protein and produce a β-strand. These results are experimental proof that the environment is important in determining the secondary structure formed by an amino acid sequence; therefore schemes that predict secondary structure from amino acid sequence alone can never be totally successful.

INTRODUCTION

Cracking the second half of the genetic code, as the protein folding problem is sometimes described, is considered one of the most important questions to be answered by researchers interested in the structure of biological molecules. It is generally accepted that the amino acid sequence of a protein determines its ultimate three-dimensional structure. In the hierarchic view, the primary structure determines regular repeating secondary structures, which in turn fold up into a tertiary structure. Researchers have noted that certain amino acids have a preference for a given secondary structure, and a number of schemes have been developed that use amino acid preferences to predict secondary structure from primary structure. The most popular methods are from Chou and Fasman (1974a,b and 1978a,b); Burgess *et al.* (1974); Lim (1974a,b); Robson and Suzuki (1976); Garnier *et al.* (1978); and Gibrat *et al.* (1987). There are also predictions of secondary structure based on homology (Pongor and Szaley, 1985, Sweet 1986, Nishikawa and Ooi 1986, Levin *et al.* 1986, Zvelebil 1987), linear optimization of predictors (Edelman and White, 1989) and neural networks (Qian and Sejnowsky, 1988; Holley and Karplus, 1989). The results of methods for predicting secondary structure from amino acid sequence were initially impressive, but have failed to improve substantially; generally, about 60% of residues can be classed correctly as α-helix, β-strand, β-turn, or other (Nishikawa and Noguchi, 1991; Yada et al., 1988; Nishikawa, 1983; Kabash and Sander, 1983; Popoz, 1980).

If the amino acid preferences were absolute, then the protein folding

problem would undoubtedly be solved. Since preferences are not absolute, one can view the protein folding problem in reverse and ask the question: Why can each amino acid be found in every type of secondary structure? If this question is answered, we might be well on our way to solving the protein folding problem, and our research here deals with this question.

In our earlier work, we have carried out CD studies that demonstrate changes in secondary structure with a change in environment, such as ligand binding, solvent variations, and quaternary interaction (Manavalan *et al.*, 1985; Manavalan *et al.*, 1986; Hennessey *et al.*, 1987; Hennessey *et al.*, 1982). The change in secondary structure with quaternary interaction is particularly interesting. The 12 S_H subunit of transcarboxylase will dissociate into 2.5 S_H monomers at pH 9.0. We found that when these monomers form an interacting hexmeric 12 S_H subunit at pH 5.8, there is a 19% increase in α-helix accompanied by a loss of 8% β-sheet and 7% other (Hennessey et al., 1982). One aspect of such interaction is to change the effective solvent that a peptide sequence sees, and the secondary structure preference of a peptide sequence may well depend upon the environment.

We have investigated three equivocal amino acid sequences that are predicted to be in one secondary structure from amino acid preferences, but are actually found in another secondary structure. These are the interesting sequences, because they are the demonstrated failures of our prediction methods. For all three equivocal amino sequences we recover the predicted secondary structure in some solvent systems. We then follow the secondary structure as a function of the solvent, ultimately mimicking the environment

inside a protein and producing the observed secondary structure. Our research demonstrates experimentally that the environment is important in determining the secondary structure formed by an amino acid sequence.

MATERIALS AND METHODS

Choosing the Peptide Sequence

Equivocal Peptides. Our first choice of an equivocal amino acid sequence came from earlier work on Eco.RI endonuclease (Manavalan *et al.*, 1984). We predicted residues 103-115 (ERE) to be in an α-helix by several primary sequence prediction methods, but that sequence forms a β-strand in the protein (McClarin *et al.*, 1986, Kim et al., 1990). To obtain more equivocal peptides we applied the Chou and Fasman method (1974a,b and 1978a,b) to the Kabsch and Sander data base (1983). Some sequences that were predicted to be an α-helix were shown primarily as β-strand in the data base. We chose the two sequences that were predicted to be α-helices about 15 amino acids long rather than shorter ones, residues 77-90 from γ-chymotrypsin (CMT) and 5-19 from apo-liver alcohol dehydrogenase (ADH). The sequences we synthesized (Table 2.1) are the three predicted to be α-helix. The portions were found to be β-strand from inspection of their X-ray diffraction (McClarin et al., 1986; Kim et al., 1990; Cohen et al., 1981; Colonna-Cesari, 1986). In order to avoid end effects we blocked the N-terminal with acetyl and the C-terminal with amine. Position 77 in CMT was changed to Trp so that all three peptides would have an aromatic residue for UV detection at 280 nm. This change did not alter the potential of the peptide for α or β structure. We also synthesized a peptide sequence CMT p11 from γ-chymotrypsin (residues 80-90), which is three amino acids shorter than CMT from the N-terminal. The whole sequence is β-strand in the parent protein. We studied it in different

solvent systems and found that it has very similar CD spectra to CMT, especially the α-helix CD. Therefore we will primarily discuss ERE, CMT and ADH.

Bioactive Peptide Somatostatin. In order to test the solvent systems we chose somatostatin, a growth hormone release inhibiting factor 14 amino acids long. It has potential for forming both α-helix in residues 6-11 and β-strand in residues 6-12 according to the Chou and Fasman method (Table 2.2), and is available commercially.

A Random Structure Sequence. We also chose a sequence from HIV-1 Tat protein (HIT) to study the solvent effect on a control peptide in the same solvent systems as we studied for the equivocal peptides. This sequence is from residues 2-23, and is predicted to be random structure by the Chou and Fasman method (Table 2.3).

Peptide Synthesis and Purification

Peptide Synthesis. The three equivocal peptides were synthesized by solid-phase methods on an automated Applied Biosystems peptide synthesizer Model 430A (University of Oregon) or 431A (The Central Service Laboratory at Oregon State University). Fast Moc chemistry is used in the 431A. This chemistry consists of a-N(9-fluorenylmethoxycarbonyl) amino acid derivatives, HBTU activation [2-(1H-benzotriazol-1-yl)-1,1,3,3,-tetramethyl-uroniumhexafluorophosphate], piperidine deprotection and a N-methylpyrro-lidone solvation system. A variety of loaded and unloaded resins are available for Fmoc chemistry. Synthesis is from the C-terminal to the N-terminal. In a

typical synthesis, the C-terminal amino acid of the peptide is loaded onto an HMP resin (p-alkoxybenzyl alcohol polystyrene) using HOBT activation (1-hydroxybenzotriazole). Each amino acid in turn is coupled to the growing peptide using HBTU activation with an efficiency of 97-99%, and standard protecting groups are used on reactive sidechains: Glu, (OtBu); Arg, (Pmc); Lys, (Boc); Gln, His, and Cys, (Trt).

The peptides were deprotected and removed from the resin using 95% TFE with one or more scavengers (H_2O, thioanisole, ethanedithiol, or phenol) for 1.5-3 hours at ambient temperature. The CSL staff will cleave, deprotect and characterize the peptide by a post-run ninhydrin analysis of the resin samples, and by peptide sequencing and/or amino acid analysis of the cleaved peptide.

Peptide HIT was donated by Dr. Erwann Loret, a postdoctoral fellow in our laboratory. Details of the peptide synthesis are given in Loret et al. (1991). Briefly, it was assembled according to the method of Barany and Merrifield (1980) on 4-(oxy-methyl)phenylacetamidomethyl (PAM) resin (0.5 mmol) (Applied Biosystems Inc., Forster City, CA) on a semiautomated synthesizer (NPS4000, Neosystem, Strasbourg, France). Side-chain protection of the butyloxy-carbonyl amino acids (Boc amino acid) (Neosystem) was as follows: Asp and Glu, cyclohexyl; Ser and Thr, benzyl; Lys, 2-chlorobenzyloxycarbonyl; Tyr, bromobenzyloxycarbonyl; Arg, tosyl; His, benzyloxycarbonyl; Trp, formyl; Cys, acetamidomethyl. The synthesis cycle used for each Boc amino acid incorporation is according to the method of Coste et al. (1990). The peptide was deprotected and removed from the resin

Table 2.1 Amino acid sequences of three equivocal peptides. they are predicted to be α-helical, but are observed with the underlined portion as β-strand.

Sequence	Residues	Notation
acetyl E W R V V L V A E A K H Q amide	103-115	ERE p13
acetyl W E K I Q K L K I A K V F K amide	77-99	CMT p14
acetyl K V I K C L A A V L W E E K K amide	5-19	ADH p15

Table 2.2 Peptide sequence of somatostatin. Solid underline: predicted helix; dashed underline: predicted β-strand.

Sequence	Name
A G C K N F F W K T F T S C	Somatostatin

Table 2.3 Peptide sequence HIT from HIV-1 Tat protein.

Sequence	Residues	Notation
E P V D P R L E P W K H P G S Q P K T A C T	(2-23)	HIT

with a high hydrogen fluoride (HF) procedure using 10% by volume of p-cresol as a scavenger.

Peptide Purification. The crude peptides were purified by high-pressure liquid chromatography (HPLC) using a VYDAC C-18 reverse phase column on an apparatus composed of the following elements: a single or double LKB HPLC pump 2150 with a solvent programmer, a V4 ISCO UV absorbance detector recording from 215-230 nm and a Beckman 163 variable wavelength detector monitored at 280 nm. These were interfaced with a IBM-compatible computer.

The solvent system was: buffer A, 0.1% TFA in water; buffer B, 0.06% TFA in acetonitrile. The usual gradient is summarized in Table 2.4a. The collected samples were dried in a speed vac concentrator (Savant) overnight or for a few hours at 30 °C. The samples were stored at -20 °C. Before using them for CD measurements, the samples were monitored by C18 HPLC. Peptide identity was verified by amino acid analysis and mass spectroscopy.

The crude peptide HIT was purified by medium-pressure liquid chromatography including: a model 590 solvent delivery system from Waters, a Labomat VS200 solvent programmer from Labomatic, and a UV spectro-photometer and a recorder from LKB. The preparative column was a C18 from Knauer. The solvent system was: buffer A, 0.1% TFA in water; buffer B, 0.08% TFA in acetonitrile. Following the MPLC run, the peptide fractions were monitored by an analytical C18 HPLC and homogeneous ones were pooled and lyophilized.

Amino Acid Analysis

In order to check the purity of the purified peptides, we used the methods developed in Dr. Sonia Anderson's laboratory (Malencik et al, 1990) to analyze amino acid compositions. Peptides were hydrolyzed with 6 M HCl in the vapor phase for either 24 hours at 110 °C or at 160 °C for 1 hour. Usually, 10 µg of salt-free peptide was placed in a 13 x 75-mm glass test tube which was previously pyrolyzed at 230 °C for 12 hours and dried in a Savant Speed-Vac. We then added 2 ml of 6 M HCl and placed the sample in a vacuum hydrolysis vessel. A blank tube, as well as a standard amino acid sample (5 nmol of each amino acid), was also included with each hydrolysis set. After flushing with nitrogen, a vacuum was applied (<20 mTorr) and the vessel placed in a temperature-controlled oven for the required length of time. After hydrolysis, the tubes were dried in a vacuum desiccator for 2 hours. For derivatization, 50 µl of 50 mM NaHCO, pH 8.5, was added to each tube followed by 100 µl of 4 mM dabsyl chloride (400 nmol) in acetonitrile. The tubes were capped with parafilm, stirred, and placed in a heat block at 60-70 °C for 10 min (Knecht, R., and Chang, J. Y. (1986). Then 850 µl of a 50 mM sodium phosphate pH 6.8 ethanol mixture (1:1) was added to each tube. Twenty-microliter portions were injected and analyzed.

The chromatographic separation is done on a 5µ C18 column of reverse-phase HPLC using a single LKB 2150 pump, a Labtronix automatic injection valve, an Isco 2360 gradient controller, and an Isco V4 detector equipped with a 5-mm path length, 3.5-µl volume flow cell. Precolumns (20 x 2 mm), packed with the same materials employed in the separating columns,

were used throughout. The gradient solution is: mobile phase A: 18 mM NaOAc, pH 6.6 with 4% dimethylformamide (DMF); mobile phase B: 90% Acetonitrile + 10% isopropanol. The gradient controller has a built-in delay system, since there is a certain amount of holdup volume between the gradient mixture and the column. At a flow rate of 1 ml/min, this was determined to be 1.5 min from the start of the gradient until the time the gradient actually reaches the column. Therefore, sample injection was initiated 1.5 min after the start of the gradient. This procedure makes the retention times more reproducible. Absorbance was monitored at 460 nm and band areas were determined using Dynamic Solution's baseline data acquisition system (version 3.0), which was programmed for on-line band identification and quantitation. All of the dabsylated amino acids exhibit a broad absorption band centered at 460-465 nm. Table 2.4 summarizes the gradient systems for peptide purification and amino acid analysis. The gradient varies a little for each peptide sequence.

Further, amino acid analysis was used to determine peptide concentration besides checking purity. In order to determine the concentration of our stock solution, amino acid analysis was performed by the staff of the Central Service Laboratory at Oregon State University. They operate a Beckman HPLC amino acid analyzer with an ion-exchange, post-column ninhydrin derivatization system. The protein hydrolysate is injected into the system through the autosampler. Each amino acid of the hydrolysate is resolved by ion-exchange chromatography on a Beckman 2 × 250 mm Spherogel column, then mixed with ninhydrin, heated to 130 °C and detected

by absorbance at 570 nm. The computer compares the chromatographs of unknown samples with those of amino acid standards and assigns and quantitates each amino acid. Protein hydrolysates for the ninhydrin system are prepared by hydrolysis in 6N HCl, 1% phenol at 110 °C for 72 hours. About 50-150 μg of protein is required for a good analysis on this system; samples for amino acid analysis must be homogeneous and salt free. We take into account that tryptophan and underivatized cysteine are destroyed during hydrolysis, and that proline results are not quantitative. All others are analyzed to within 10% of the true value. For a more accurate analysis, triplicate samples were used. Amino acid analysis gave us more accurate concentrations and confirmed that our samples are pure.

Solvent System

Since we are trying to mimic the interior of a protein, we used any solvent that would stabilize secondary structure. All the organic solvents used were 99+% purity or spectrophotometric grade. Different mixtures of TFE (Aldrich Chemical Company, Inc.), acetonitrile (Matheson Coleman and Bell), methanol ('Photrex'), ethanol (Midwest Grain Products Co.), octanol (Aldrich), 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, Aldrich), cylcohexane (MCB or Eastman spectro), and pentane (Aldrich) were used as solvent systems. Other solvents were also used, such as 1-o-octyl-β-d-glucopyranoside (Aldrich Chem. Co), and sucrose (Schwarz/Mann). We found that many organic solvents stabilize the α-helix for our peptide sequences. The detergent sodium dodecyl sulfate (SDS), electrophoretic grade from BioRad, was also

used to promote secondary structure in our peptides.

Spectroscopy

UV Measurements. Absorption was measured on a Cary 15 spectrophotometer, purged with nitrogen when used in the far UV range. Extinction coefficients for these peptides in buffer were established by the guanidine hydrochloride method (Elwell and Schellman, 1977). First, 3.2 ml buffer, 0.8 g Gn-HCl and 20 µl β-mercaptoethanol (if there is a S-S bond) were mixed; a 1 cm cell (600 µl volume) placed in the sample chamber and a 1 cm cell (3 ml volume) placed in the reference chamber of the spectrometer both filled with the mixture, were used to measure the baseline for absorption. Second, the sample was dissolved in 400 µl 10 mM $NaHPO_4$, pH 7.0 buffer, and 0.4 g Gn-HCl and 10 µl β-mercaptoethanol were added to the sample to give 710 µl in volume. The sample replaced the buffer in the small 1 cm cell in the sample chamber and absorption at 280 nm was measured. Third, the same concentration of sample in buffer was prepared, and the OD of this sample at 190 nm in a 0.02 cm cell was measured. Fourth, from the $OD_{280\ nm}$ and the known extinction of Trp, Tyr, and Cys at 280 nm, we can obtain the amino acid concentration of the sample. In this way, the extinction coefficient at 190 nm on a per amide basis was obtained from Beer's law. We also dilute the stock solution at a known protein concentration from the amino acid analysis discussed above to obtain extinction coefficients. The values obtained by these two methods are very close. Then small amounts of sample from the same stock solution were diluted in different solvents to

Table 2.4a Gradient for purification of three peptides.
Solvent A: 0.1% TFA in water
Solvent B: 0.06% TFA in acetonitrile
Flow rate: 0.75 ml/min
Temperature: 23 °C

Time (min)	%B
0	10
5	40
15	50
20	70
22	70
24	10

Table 2.4b Gradient for amino acid analysis.
Solvent A: 18 mM NaOAc, pH 6.6 with 4% DMF
Solvent B: 90% Acetonitrile 10% isopropanol
Flow rate: 1.0 ml/min
Temperature: 40 °C

Time (min)	%B
0	18
6	25
26	40
41	70
43	70
45	18

Table 2.5 Extinction coefficents of three equivocal peptides in buffer, TFE, and SDS.

Peptides	ε_{190} (in Buffer)	ε_{190} (in TFE)	ε_{190} (in SDS)
ERE p13	10145	9225	9560
CMT p14	11240	10990	11660
ADH p15	8420	7930	8500

obtain their extinction coefficients. The extinction coefficients for these peptides under different conditions are shown in Table 2.5. The extinction coefficient for α-helix has the lowest value while the extinction coefficient of β-sheet has the highest. This is in good agreement with the fact that an α-helix usually has a lower absorption at 190 nm while β-sheets have a higher absorbance at this wavelength (Rosenheck and Doty, 1961). However, under some conditions this is not necessarily true and our values for the three conditions are quite close, indicating that the absorption at 190 nm does not change much with different solvents.

Vacuum UV CD Spectroscopy. CD is measured by alternately passing left- and right-handed circularly polarized light through a sample of asymmetric molecules. The difference in absorption is related by Beer's Law to the difference in extinction coefficients:

$$\Delta A = A_l - A_r = (\varepsilon_l - \varepsilon_r)c\ell = \Delta\varepsilon c\ell \qquad (2.1)$$

where c refers to the concentration in moles per liter, ℓ the pathlength of the sample in centimeters, and the rotation of the light is denoted by the sub-scripts. $\Delta\varepsilon$ is a characteristic of the molecule with a sign and magnitude that is a function of wavelength. CD theory and its practical application, including the CD of proteins, can be found in many reviews. (Mason, 1982; Tinoco and Cantor, 1970; Johnson 1971; Schellman, 1975; Woody, 1981; Woody 1985; Yang et al., 1986; Johnson, 1988).

CD spectra of freshly prepared samples were taken on a McPherson

vacuum UV spectrophotometer modified for CD as described elsewhere (Johnson, 1971). Measurements were made using quartz cells of various pathlengths, usually 0.05 cm to 0.01 cm. The instrument was calibrated using (+)-10-camphorsulfonic acid, $\Delta\varepsilon$ = +2.37 $M^{-1}cm^{-1}$ at 290.5 nm and -4.95 at 192.5 nm (Chen and Yang, 1977). The results were digitized at 0.5 nm intervals using an IBM-type computer system that collected the data at a rate of 1 nm per minute. Spectra were smoothed using a cubic spline algorithm.

Near UV Spectroscopy. In order to monitor solutions for aggregation, CD spectra from 260 to 210 nm as a function of concentration were measured on a Jasco J-40 spectrometer. These experiments were carried out at 5°C, and if there was no CD change with concentration, further studies were carried out at 25°C, 45°C, and 60°C. We chose different pathlength cells ranging from 1 cm to 0.01 cm to accommodate the different concentrations.

Data Analysis

Singular Value Decomposition. One of the most widely used applications of CD is the determination of secondary structure composition of proteins in solution (see Johnson, 1988 for review). In this approach, an experimental Vacuum UV (260-178 nm) CD spectrum is analyzed for secondary structure using basis spectra derived from CD spectra of proteins having a known secondary structure determined by X-ray crystallography.

For a more detailed description of the procedure, see discussions of the singular value decomposition (SVD) approach of Johnson and co-workers

(Johnson, 1988, Hennessey and Johnson, 1981, Compton and Johnson, 1986). Briefly, Hennessey and Johnson (1981) applied SVD to the CD data for 16 reference proteins of known secondary structure. They used the five most important basis vectors from SVD to predict the secondary structure of a protein from its CD spectrum. This method was further improved as a more simple and straightforward procedure by Compton and Johnson (1986). Since CD spectra are related to a protein's secondary structure, according to this method the linear relationship between CD data and their corresponding secondary structure can be written as the matrix equation

$$\mathbf{XC} = \mathbf{F} \tag{2.2}$$

Here, **C** represents the CD data of the reference proteins, **F** is the corresponding fraction of the protein secondary structure, and **X** is the matrix that transforms CD spectra into their related secondary structure. **X** can be determined by transforming Equation 2.2 to

$$\mathbf{X} = \mathbf{F}\mathbf{C}^{-1} \tag{2.3}$$

Where $\mathbf{C}^{-1}$ is the inverse of **C**. According to the singular value decomposition theorem (Nobel and Daniel, 1977), any matrix can be decomposed into the product of three matrices:

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^{T} \tag{2.4}$$

The **U** matrix contains new CD basis vectors that are all orthogonal. The **S** matrix contains singular values on the main diagonal and zeros elsewhere. The singular values give the importance of each orthogonal CD basis vector in reconstructing the original protein CD spectra. The singular value in each column corresponds to a column in the **U** matrix. The product **US** is a matrix with columns that can be considered to be basis CD spectra. **US** has the same dimensions as the original data matrix, and the columns look like typical CD spectra when expressed graphically. The $\mathbf{V}^T$ matrix contains the coefficients which fit the basis CD spectra to the protein CD spectra in the original data matrix, **C**. **U** and $\mathbf{V}^T$ are orthogonal and unitary matrices. Then

$$\mathbf{C}^{-1} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^T \tag{2.5}$$

Where $\mathbf{S}^{+}$ has nonzero entries only on the main diagonal which are the reciprocals of the singular values (Forsythe et al., 1977). In calculating $\mathbf{C}^{-1}$, we use only the five most significant singular values and their corresponding vectors in the $\mathbf{U}^T$ and **V** matrices to eliminate noise and avoid instability. A computer program was developed to do all the complicated calculations, allowing us to use this method easily and quickly.

Once we have matrix $\mathbf{C}^{-1}$ and the secondary structures from X-ray data for the basis proteins, we can obtain the **X** matrix from Equation 2.3. Multiplying digitized CD spectra by **X** gives us the secondary structure of the unknown protein from Equation 2.2.

Variable Selection. Variable selection removes unimportant

variables from an underdetermined system of equations. For instance, there are five independent variables or equations in the CD spectrum of a protein measured into the VUV region to 178 nm. However, the CD spectrum of a protein depends upon many parameters, e.g., α-helix, parallel and antiparallel β-sheet, the various types of β-turn, aromatic side chains, prosthetic groups, nonrepetitive structures, tertiary folding etc. Because the CD spectrum of a protein is analyzed using the CD spectra of a set of reference proteins with known secondary structures, it is reasonable to remove reference proteins that have contributions to their CD spectra from parameters not found in the protein to be analyzed with the set. The goal of removing reference proteins is to eliminate factors contained in reference proteins that are not found in the protein being analyzed, so that the number of variables can be reduced to five, which is consistent with the information content of our data (Manavalan and Johnson, 1987).

If there are n proteins in the basis set and r proteins to be removed, there are (n! - (n-r)!)/r! combinations. The secondary structure is recalculated for all possible subsets. Here secondary structural analysis was performed on all possible combinations of r=3 using a 26 protein basis set (Toumadje and Johnson, 1992). The protein whose provisional removal leads to the most combinations giving an improvement in both secondary structural analysis and curve-fit is completely removed from the basis set and the procedure is repeated on n-1 proteins until the following criteria are met: (1) The sum of secondary structures is in the range of 0.9 and 1.1, or 0.95-1.05, if possible. (2) No negative values greater than -0.05 are obtained from any particular

secondary structure. (3) The reconstructed spectrum is within experimental error. This is equivalent to the root-mean-square of error being less than 0.22 $\Delta\varepsilon$. (4) The basis set having more reference proteins is preferred. (5) Proteins with CD spectra that closely resemble the CD of the protein being analyzed are included in the basis set. Many subsets will give similar results, and subsets that satisfy all the criteria can be averaged to give the predicted secondary structure for the protein. Recently we modified our criteria for acceptable combinations as follows (Toumadje and Johnson, 1992), with priority given to the sum of secondary structures:

(1) Analyze the CD with all 26 proteins in the basis set to create only one combination. If there are no negative values, we exclude all combinations with negative values in our acceptable combinations in subsequent variable selection; otherwise no negative values greater than -0.05 are allowed.

(2) Run variable selection, removing various combinations of proteins from the basis set and choose the combinations with a total secondary structure between 0.96-1.05, regardless of the fit of the reconstructed CD. If there are many such combinations, narrow the range for the total and choose the ones closest to 1.00 so that there are about twenty combinations. Then begin improving the fit until there are about ten combinations having a total close to 1.00 with the best fit.

The vacuum UV CD spectra from 260 nm to 178 nm were analyzed using singular value decomposition combined with variable selection as described above. We used a 26 protein basis set which contains many spectra of all-β proteins. Our percentages of secondary structure are given on

a per amide basis. Since our CD secondary structure predicting program is based on different combinations of secondary structures in proteins, it does not work well for the single secondary structure induced in our peptides, especially for β-sheet. Therefore we also interpreted secondary structure in our CD spectra by comparing it with published CD spectra for pure secondary structures (Greenfield and Fasman, 1969). Furthermore, $\Delta\varepsilon_{222\,nm} \times (-10)$ is a good estimate of the percentage of α-helix, as Figure 2.1 shows. There is a simple linear relationship between α-helix content and ε; for example, if an unknown protein shows a Δε of –6 at 222 nm, its percentage of α-helix will be about 60%.

Figure 2.1 CD at 222 nm for 26 proteins as a function of their α-helix content from X-ray studies.

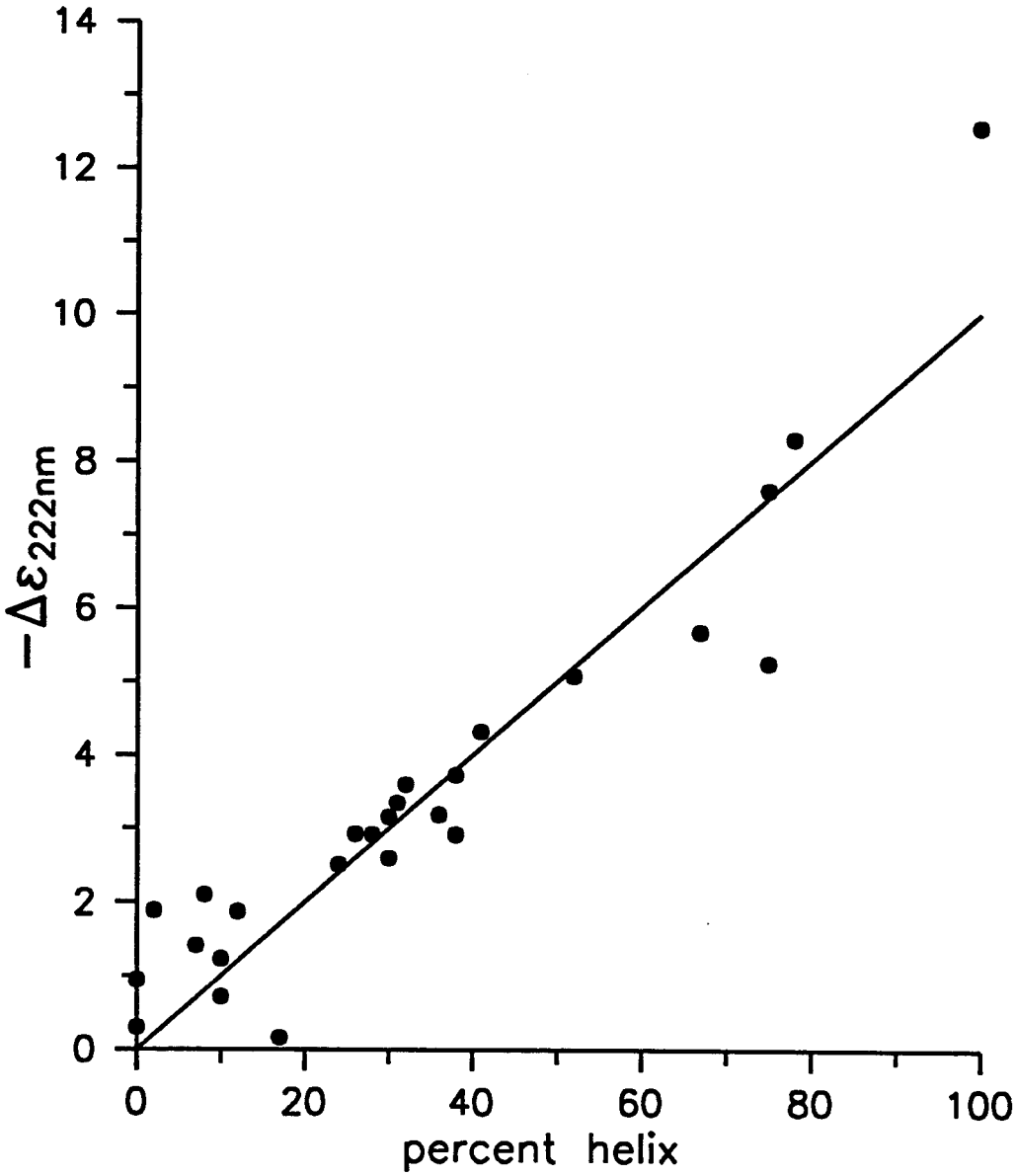
14
12
10
8
6
4
2
0
$-\Delta\varepsilon_{222nm}$
0
20
40
60
80
100
percent helix

Figure 2.1

RESULTS AND DISCUSSION

Solvent Effects on a Bioactive Peptide

To begin our research on equivocal peptides, we repeated the work of Wu and Yang (1981) and Wu et al. (1981) on somatostatin that has been reduced by dithiothreitol. Somatostatin is a growth hormone release inhibiting factor 14 amino acids long. Native somatostatin has a disulfide bond between Cys-4 and Cys-14. Yang and coworkers found that this polypeptide has a helix-forming potential for residues 6-11, but the helical conformation is constrained by the presence of a single disulfide bond. Once the S-S- linkage is broken, this peptide becomes partially helical in 25 mM SDS. According to the Chou-Fasman method, somatostatin also has a β-strand forming potential for residues 6-12. Thus, in a low SDS solution (2 mM) it adopts a β-strand structure rather than an α-helix. The reduced somatostatin is a random coil in aqueous solution. Somatostatin-surfactant interactions were also studied by Holladay and Wilder (1980). Their CD and electron paramagnetic resonance studies of somatostatin reveal that SDS markedly alters the conformation of the peptide, whereas Triton X-100 and Lubrol WX have no effect, and dodecyltrimethyl- ammonium chloride only a slight effect.

We repeated Wu and Yang's work and extended the spectra to 185 nm. Our intensities are somewhat more accurate, because we used amino acid analysis for determining concentration, rather than weighing the materials out of the bottle. Our analysis showed 32% α-helix in 25 mM SDS and 35% β-strand in 2.2 mM SDS (Table 2.6), these fractions being about 70% of the

values predicted by the Chou and Fasman method (43% α and 50% β). The spectrum of reduced somatostatin in water superimposes on that of the native peptide. The presence of an α-helix in solution is easily identified by the CD spectrum (Figure 2.2) measured between 185-260 nm, due to a well-characterized positive band at 190 nm, a negative band at 205-207 nm related to the π-π^* transition, and a negative band at 215 nm related to the n-π^* transition (Johnson, 1985; Woody 1985). Somatostatin as an α-helix has a double minimum at 225 nm and 207 nm, and a positive band at 195 nm while somatostain as a β-strand has a negative band around 214 nm and a positive band at 198 nm. Cationic side groups are believed to be essential for the formation of a β-structure in SDS (Wu and Yang, 1980), and there are two Lys residues in somatostatin neutralized by SDS. Wu et al. (1981) also lowered the pH of the solution from 6.9 to 2.1. The CD of α-helix and β-strand did not change because somatostatin has no Glu or Asp residues.

Surfactants are used extensively in the biochemical study of macromolecules. For example, the anionic SDS is used in gel electrophoresis to determine the molecular weights of proteins (Weber and Osborn, 1969), while the nonionic Triton X-100 and Lubrol WX are used to solubilize membrane proteins (Helenius and Simons, 1975). Surfactants are considered as denaturants but how they work is not well understood. Urea or guanidine hydrochloride unfold proteins and destroy the ordered structure at a concentration above 5 M, while surfactants such as SDS, destroying the tertiary structure but keeping the secondary structure, can alter protein

Figure 2.2 CD of somatostatin in SDS solutions. As a random coil in sodium phosphate buffer, pH 7.0 (.....), as an α-helix in 25 mM SDS (-----), and as a β-strand in 2 mM SDS (- - - -).

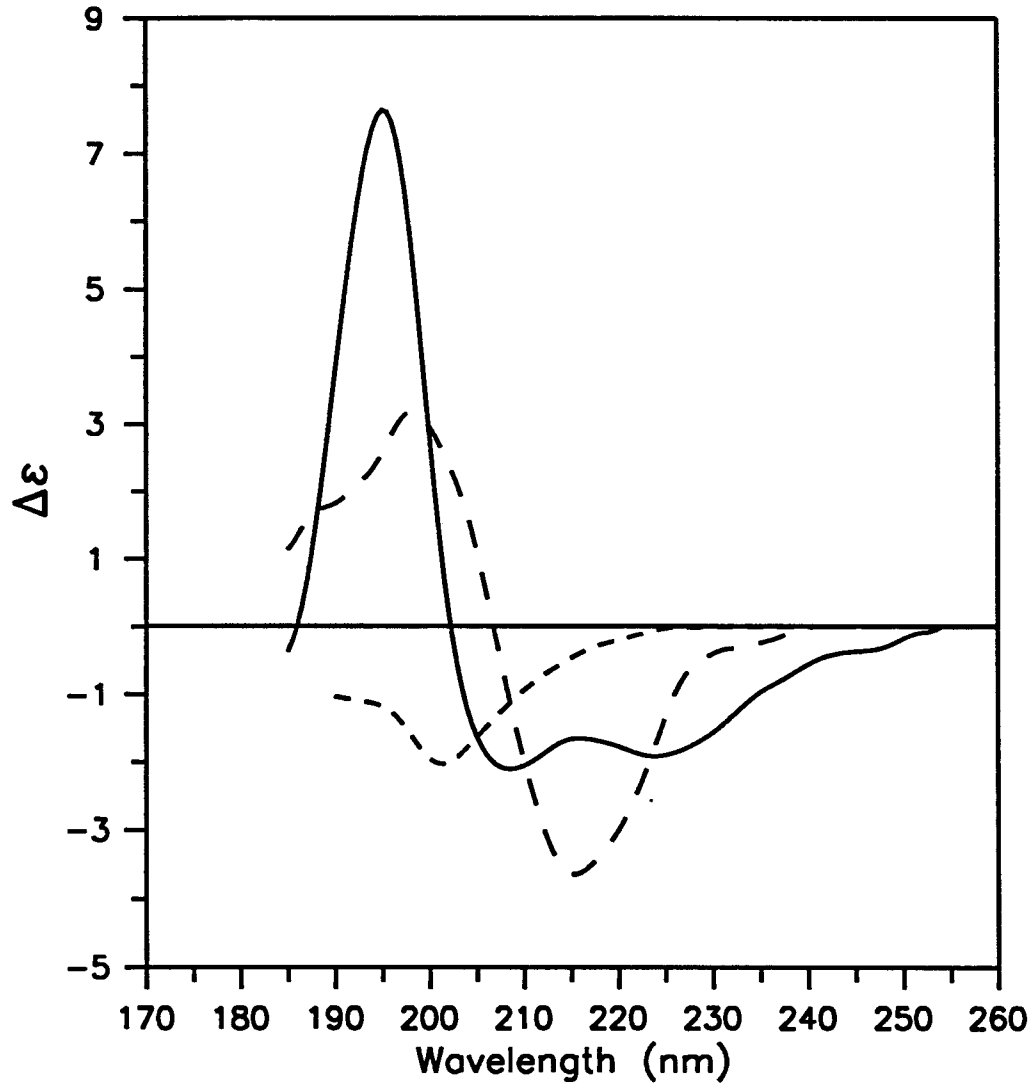
9
7
5
3
1
−1
−3
−5
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.2

conformation at 1 mM or less (Tanford, 1968). At these low concentrations SDS can either increase or reduce the helical structure in proteins, and can convert proteins rich in β-conformation to helical ones (Mattice et al., 1976). Many proteins in aqueous solution can change their conformation with SDS (Reynolds and Tanford, 1970; Visser and Blout, 1971; Jirgensons, 1976; Mattice et al., 1976; Su and Jirgensons, 1977). SDS micelles can induce helix folding in some proteins (Kubota et al., 1987; Anel et al., 1991), while some proteins rich in β-form can be converted into helices in excess SDS solutions, such as concanavalin A (Kay, 1970) and elastase (Visser and Blout, 1971).

Yang and coworkers chose commercially available oligopeptides and short polypeptides for their studies in surfactants because of their structural simplicity. They reported that the induced conformation of oligo- and polypeptides in surfactant solutions depends on their amino acid sequence, which in turn dictates their structure-forming potential (Wu and Yang, 1978; Yang and Wu, 1978). In a SDS solution the anionic surfactant first binds to the cationic side groups of polypeptides. Additional surfactant ions then cluster around the polypeptide chain, segments of which are induced to adopt an ordered structure. At low molar surfactant/peptide ratio a β-form can exist if a segment has the β-forming potential, but excess surfactant usually disrupts the β form, and may convert it into a helix if the segment also has helix-forming potential. Polypeptides without any structure-forming potential remain unordered regardless of the surfactant concentration used.

We have investigated the effect of various organic solvents on the secondary structure of reduced somatostatin. The α-helix spectrum in 90% TFE displays a double minima at 222 nm and 208 nm, and a maximum at 190 nm (Figure 2.3). There are crossovers at 208 and 185 nm. The β spectrum in 90% CH_3CN is characterized by a negative band at 205 nm and a positive band at 190 nm, which is blue-shifted from the β-strand in SDS solution. The CD in 90% TFE analyzes to give 51% α-helix while the CD in 90% acetonitrile analyzes to give 52% β-strand (Table 2.6). This is a very good agreement with the prediction from the Chou and Fasman method, with 6 out of 14 amino acids having α-helix potential and 7 out of 14 amino acids having β forming structures. However, in SDS solution only about 30% α-helix and β-strand form in this peptide. Clearly, our experiments demonstrate that environment can affect the secondary structure of an amino acid sequence. Here both detergent and organic solvent can stabilize secondary structure of somatostatin, but the organic solvent is more effective.

Recovering the Predicted Secondary Structure

ERE p13, CMT p14 and ADH p15 are all predicted to be an α-helix, but found to be β-strand in their respective proteins. The isolated peptides are randomly structured in aqueous solution based on their CD spectra (Figure 2.4a,b,c). However, we have been able to find a variety of solvents that will recover the predicted α-helix secondary structure. Methanol, ethanol, acetonitrile, 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) with low pH, a mixture of octanol and other alcoholic solvents (such as cyclohexane, pentane, ethanol),

Figure 2.3 CD of somatostatin in organic solvents. As an α-helix in 90% TFE (——), and a β-strand in 90% CH_3CN (- - - -).

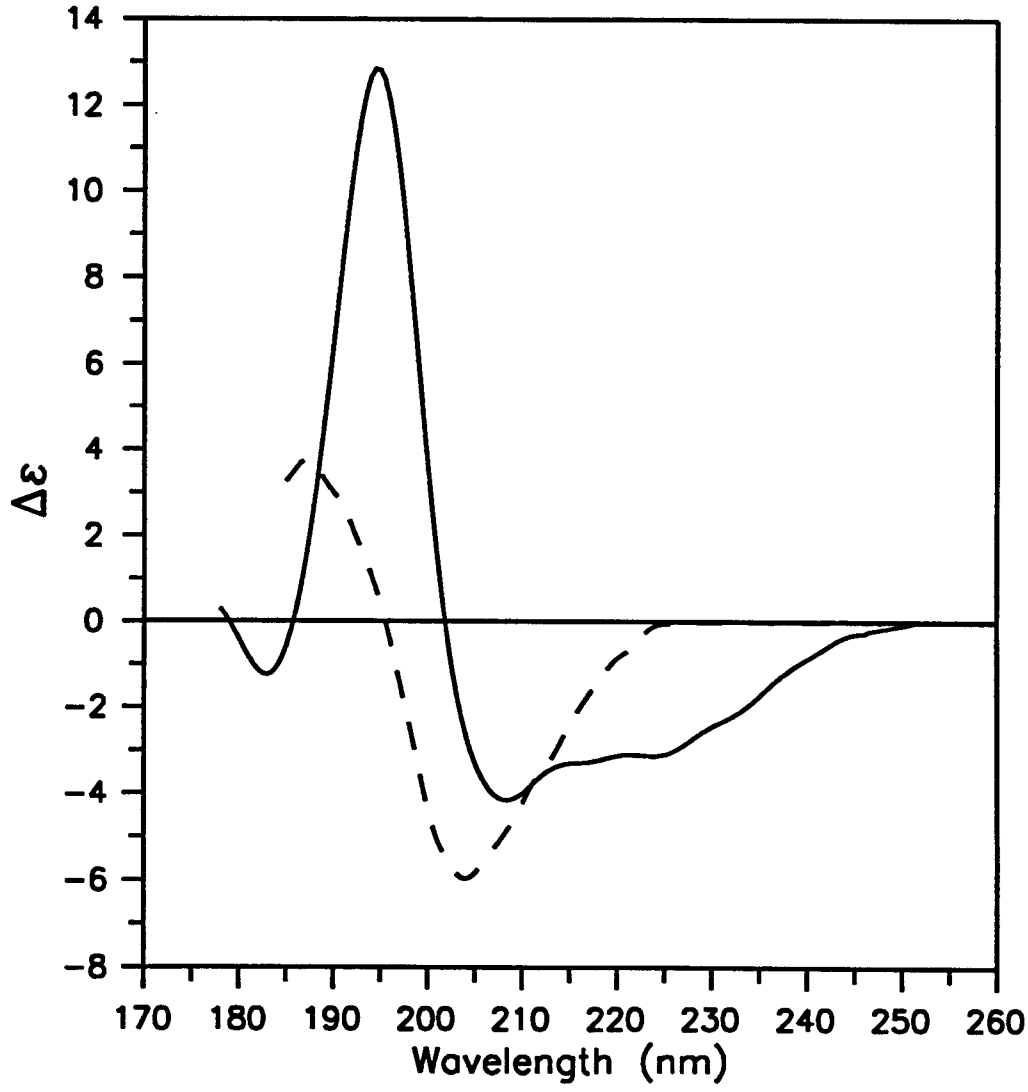
14
12
10
8
6
4
2
0
-2
-4
-6
-8
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.3

Table 2.6 Secondary structure analysis of somatostatin in different solvents.

Solvent	H	A+P	T	O	Total
25 mM SDS	0.32±0.05	0.22±0.12	0.26±0.06	0.17±0.03	0.97
2.2 mM SDS	0.20±0.01	0.31±0.01	0.00±0.01	0.53±0.02	1.04
90% TFE	0.51±0.01	0.16±0.03	0.31±0.02	0.02±0.02	1.00
90% CH_3CN	0.24±0.02	0.51±0.05	0.28±0.02	0.00±0.01	1.03

TFE, and high concentrations of SDS (about 25 mM) all give a typical α-helical CD. Among them 100% TFE at 0°C gives the largest percentage of α-helix, about 77% for ERE p13, 65% for CMT p14, and 66% for ADH p15 from our analysis (Table 2.7). The analysis also yields 15% to 26% β-turn for these three peptides, without β-strand or other structure. This β-turn could be a 3_{10}-helix. If the $\Delta\varepsilon$ at 222 nm is used to monitor the percentage of α-helixfor these peptides, values of 80%, 65% and 60% are also obtained. All the α-helix spectra for the three peptides are characterized by a double minimum at 208 and 222 nm and a maximum at 190 nm (Figure 2.4a,b,c). However, in the spectrum of ADH and CMT, the negative band at 208 nm is much larger than the band at 222 nm compared to ERE. In 50% TFE, a very high percentage of α-helix is obtained, more than 70%. If TFE is increased to 90%, the percentage of α-helix is almost the same as in 100% TFE. The nature of the transition induced by TFE has also been examined. We have studied some of these sequences in 100% TFE at 0 °C, 10 °C, 25 °C, 45 °C over the range from 260 to 178 nm. By comparing the shapes of the curves (not shown) and the secondary structure data from SVD analysis, it is clear that as the temperature is increased, the percentage of α-helix decreases while the percentage of random structure increases; this indicates that a structural transition from α-helix to random-coil is taking place. The isosbestic point at 202 nm is consistent with the idea that each amino acid is involved in a two-state transition, meaning that a residue is either in an α-helical or random coil state. We also studied some of these equivocal peptides in 90% ethanol at 0 °C, 10 °C, 23 °C and 45 °C. For example, at room temperature

Figure 2.4a CD of the equivocal amino acid sequence ERE in different solvent systems. As a random coil in 10mM sodium phosphate buffer, pH 7.0 (.....), as an α-helix in 100% TFE (-----), and as a β-strand in 3.3 mM SDS (- - - -).

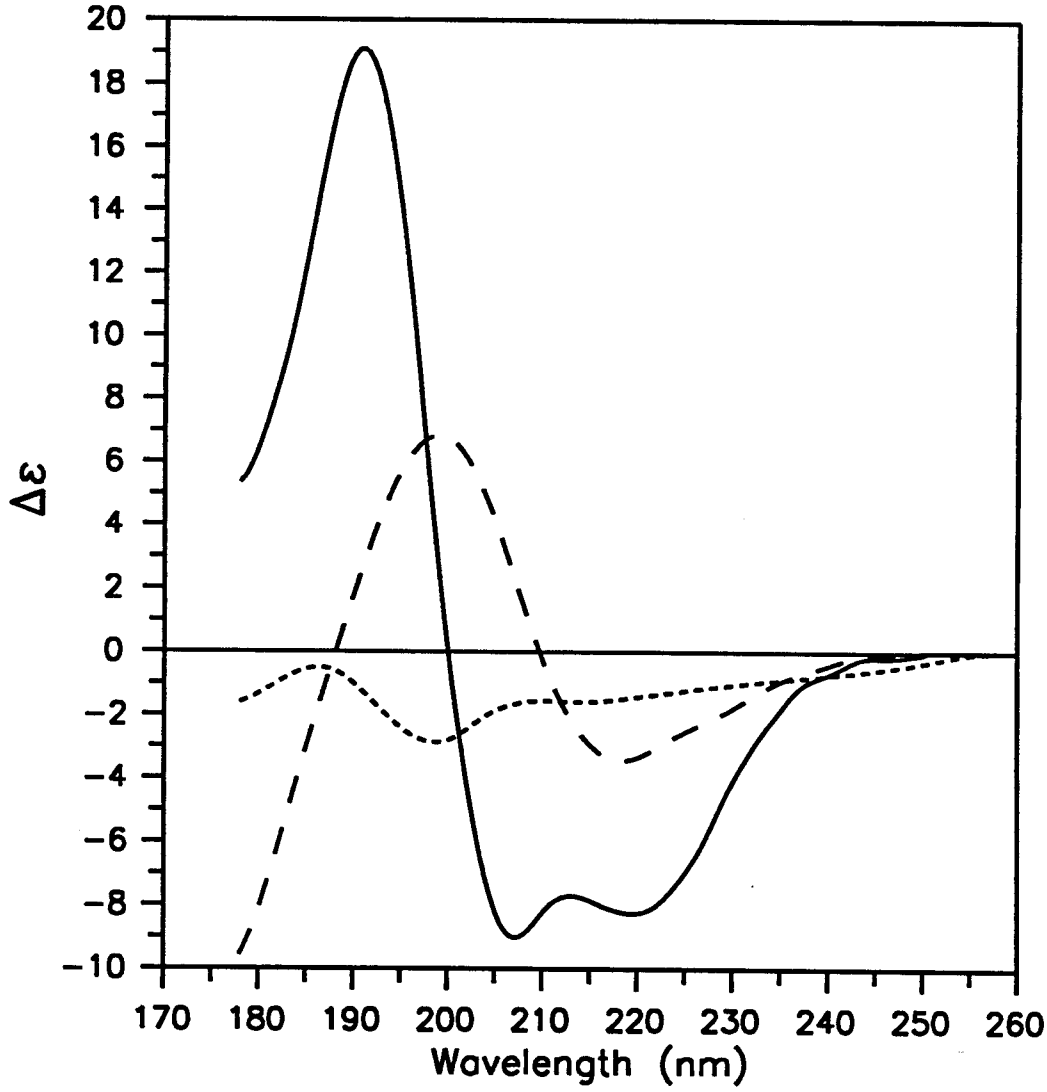
20
18
16
14
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.4a

Figure 2.4b CD of the equivocal amino acid sequence CMT in different solvent systems. As a random coil in 10mM sodium phosphate buffer, pH 7.0 (.....), as an α-helix in 100% TFE (-----), and as a β-strand in 3.5 mM SDS (- - - -).

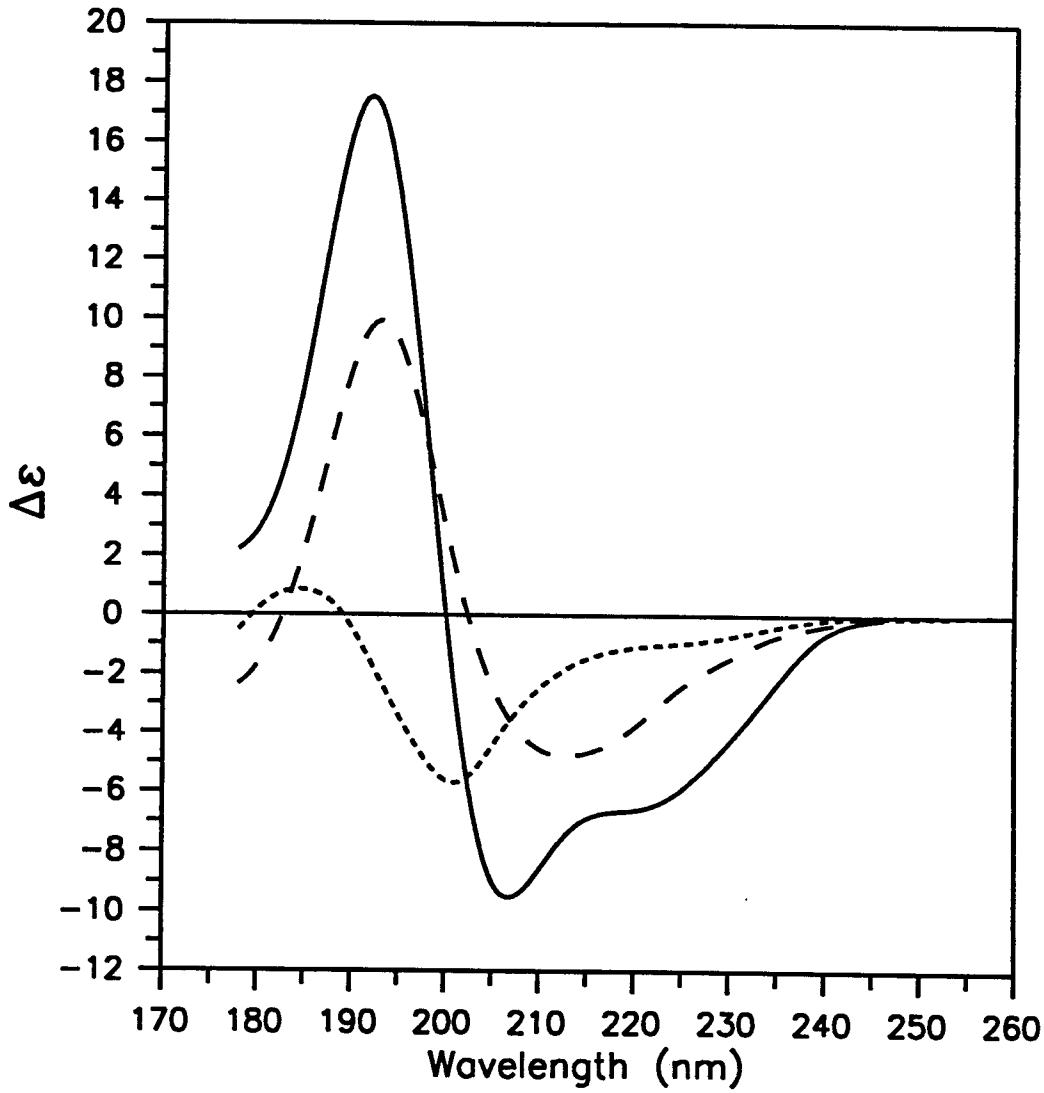
20
18
16
14
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
−12
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.4b

Figure 2.4c CD of the equivocal amino acid sequence ADH in different solvent systems. As a random coil in 10mM sodium phosphate buffer, pH 7.0 (.....), as an α-helix in 100% TFE (-----), and as a β-strand in 2 mM SDS (- - - -).

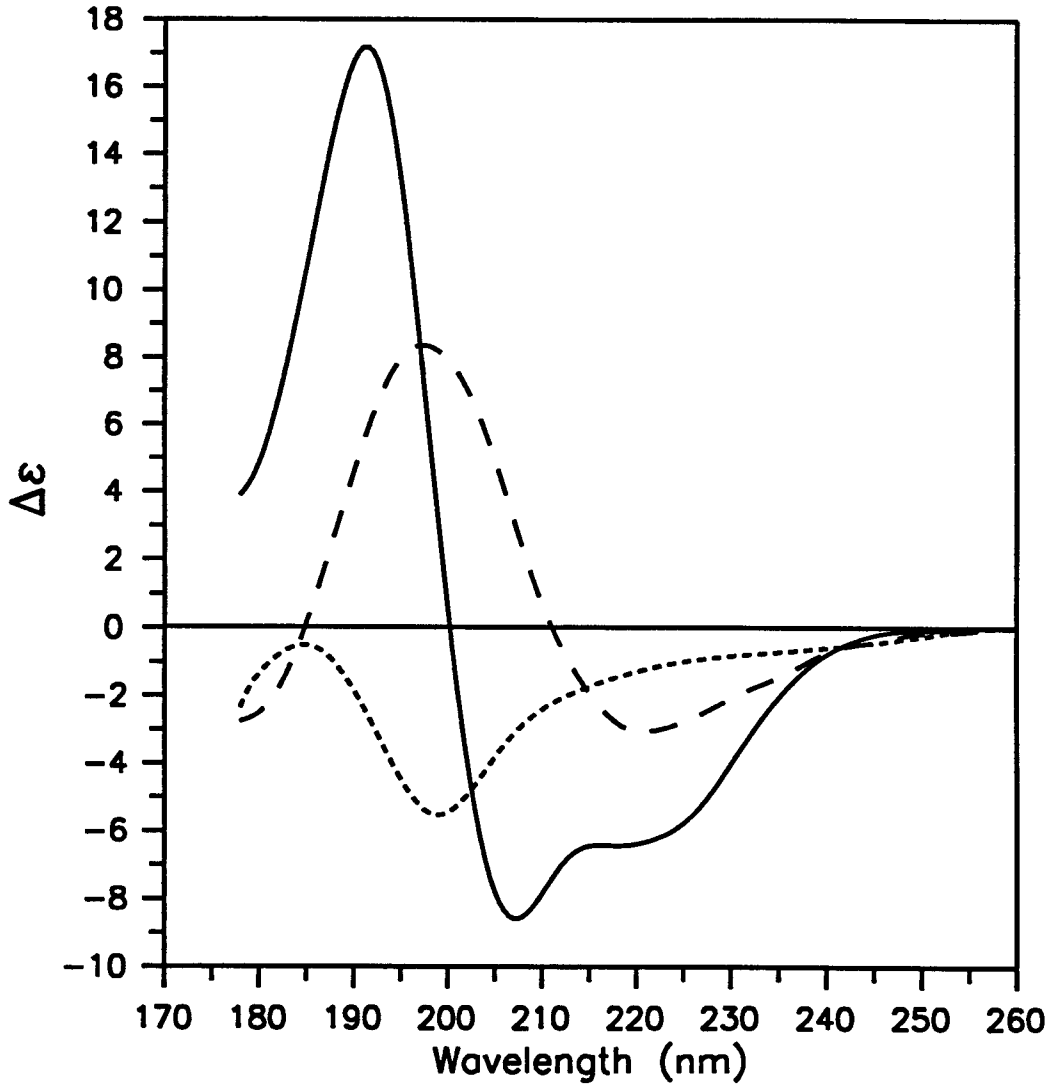
18
16
14
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.4c

Table 2.7 Analysis for secondary structure of the CD of equivocal peptides in 100% TFE.

Peptide	H	A	P	T	O	Total
ERE p13	0.77±0.01	0.04±0.03	0.04±0.01	0.15±0.01	0.03±0.01	1.03
CMT p14	0.65±0.04	0.04±0.04	0.12±0.06	0.21±0.07	0.02±0.02	1.03
ADH p15	0.66±0.03	0.05±0.04	0.02±0.03	0.26±0.03	0.00±0.03	0.99

about 50% α-helix of CMT p11 is obtained. The amount of α-helix increases to 60% at 0 °C but decreases to 18% at 45 °C. The CD spectra also show an isosbestic point at 202 nm. These results also suggest that the peptide sequence in ethanol is in dynamic equilibrium between α-helix and random structure and the transition is a two-state system.

The mechanism of helix stabilization by TFE remains unknown. However, some physical constants of TFE can be presented here: The dielectric constant of TFE is 26.67, which is about one-third that of water, 78.54, at 25 °C. Interactions between charged species would be expected to be stronger in TFE solutions. TFE is a much weaker base with a pKa_1 of about –8.2 compared to water with a pKa_1 of about -1.8; and a stronger acid with a pKa_2 of 12.4 vs. 15.3 for water (Llinas and Klein, 1975). Thus TFE is stronger at donating protons for hydrogen bonds but weaker at accepting protons in such bonds.

Llinas and Klein (1975) used ^1H NMR to study solvation effects on electron density distribution in the amides of peptides. Hydrogen bonding from the peptide amide proton (NH) to a solvent acceptor, and from a solvent donor to the amide carbonyl (C=O) decreased electronic shielding of the proton, causing a chemical shift to lower field. When the NH of an amide is exposed to solvent but the C=O is buried, the chemical shift of the amide proton moves to higher field when changing from water to TFE, because of reduced hydrogen bonding from the NH to the less basic TFE. In contrast, when the C=O is exposed but the NH is buried, the results are opposite. When both NH and C=O are exposed, the reduced hydrogen bonding of TFE to NH due

to basicity predominates.

These properties of TFE might change some interactions of the equivocal amino acid sequences in solution, and are responsible for helix stabilization of these sequences. As TFE concentrations increase, charge-charge interactions might be expected to be strengthened because of lowering of the dielectric constant. Hydrogen bonding should become more important also. The changes in hydrogen bonds between the amide proton to the carbonyl in the α-helix and the solvent would affect the extent of stabilization of α-helix. Hydrophobic interactions might also be changed, most possibly disrupted to some extent (von Hippel, 1965; Brandts and Hunt, 1967; von Hippel and Schleich, 1969). Other effects could occur, such as binding of TFE to the peptide.

Nelson and Kallenbach (1986) made a detailed study of the influence of charged groups in the ribonuclease S-peptide on its helix stabilization in TFE. They carried out CD measurements of S-peptide in TFE as a function of pH by changing either the TFE concentration or the temperature. They found that TFE does not significantly alter the magnitude of the stabilization contributed by the charged groups. In contrast, aqueous solution studies on the S-peptide showed that the charged-group effect was predominant. The observation that TFE affects interactions other than those of charged groups might indicate that many interactions are important for α-helix stability. Here we did not study the charged group effect of these equivocal amino acid sequences in TFE in detail, but we expect that the α-helix is stabilized by many interactions in addition to the effect of charges.

TFE has a reputation for promoting α-helix (Urry et al., 1971; Nelson and Kallenbach, 1986, 1989; Merutka and Stellwagen, 1989). Nevertheless, there are many reports of stable β-strands in TFE (Goodman et al., 1970; Balcerski et al., 1976; Kelly et al., 1977; Narayanan, 1986). To mention a few, peptides containing the hydrophobic side-chains Ala, Val, Leu, and Ile were synthesized as Boc-(amino acid)$_n$-OMe, where for $n > 6$, a β-strand conformation was favored in TFE. Also, the C terminal 13-mer of the NF-M subunit of human neurofilaments can form a β-strand in TFE when Ca^{2+} is present (Hollosit, et al., 1992).

TFE is often used in studies of peptide structure. For example in TFE, the helix charge-dipole effect and the helix stop signal remain operative (Nelson and Kallenbach, 1986, 1989). TFE causes artificial coiled-coiled peptides to dissociate, but retain high helical content (Lau et al., 1984). β-strands aggregate and four helix bundles are associated by TFE (Mutter and Hersperger, 1990, Epand and Scheraga, 1968). A signal sequence can form an α-helix in TFE (Bruch et al., 1989). Initial structures previously detected in aqueous solution have been shown to stabilize in TFE by NMR experiments (Dyson et al., 1988). However, to date the effect of TFE on interaction in proteins is not well understood. Prediction of the effects of TFE on a specific peptide is still difficult, because the interactions that determine peptide structure are complicated (Lehrman et al., 1990). We believe TFE may mimic important hydrophobic interactions in the interior of a protein, disrupting non-native ones when short peptides are in water. In general, TFE is a hydrophilic and hydrogen bonding solvent that stabilizes peptides in the structure

expected from the amino acid preferences used to predict secondary structure. It appears to stabilize the secondary structure for which a sequence has propensity.

To check whether the helical structure in each peptide is intramolecular or intermolecular, the CD spectra of these peptides were measured versus concentration. In Figure 2.5a, the dependence of $\Delta\varepsilon$ at 222 nm on concentration is shown for the three peptides at 5 °C. No such dependence is observed over a 200-fold range for ERE p13 and CMT p14. However, for ADH p15 the CD spectra changed with concentration, demonstrating that aggregation is part of the solvent system stabilizing the α-helix. Further studies to test for tight intermolecular interaction are shown in Figure 2.5b. As the helix in ERE p13 and CMT p14 unfolds with increasing temperature, the presence of an initial tight association at low temperature should be revealed as a dependence of the unfolding on peptide concentration (Ho and Degrado, 1987). Neither ERE nor CMT (not shown) show CD that depended on peptide concentration at any point in the unfolding, demonstrating that helix formation by these two peptides is a monomolecular process, and is not the result of aggregation. In fact, some peptides which have been studied in TFE do not aggregate. For instance, Bruch et al. showed that the LamB signal peptide in TFE is monomeric in the range of concentrations from 0.4 uM to 2 mM, and no aggregation of ribonuclease S-peptide in TFE was detected by Nelson and Kallenbach (1986).

Figure 2.5a CD of three equivocal amino acid sequences at 222 nm in TFE as a function of peptide concentration at 5 °C. Peptides ERE (-O-), CMT (-□-), and ADH (-●-).

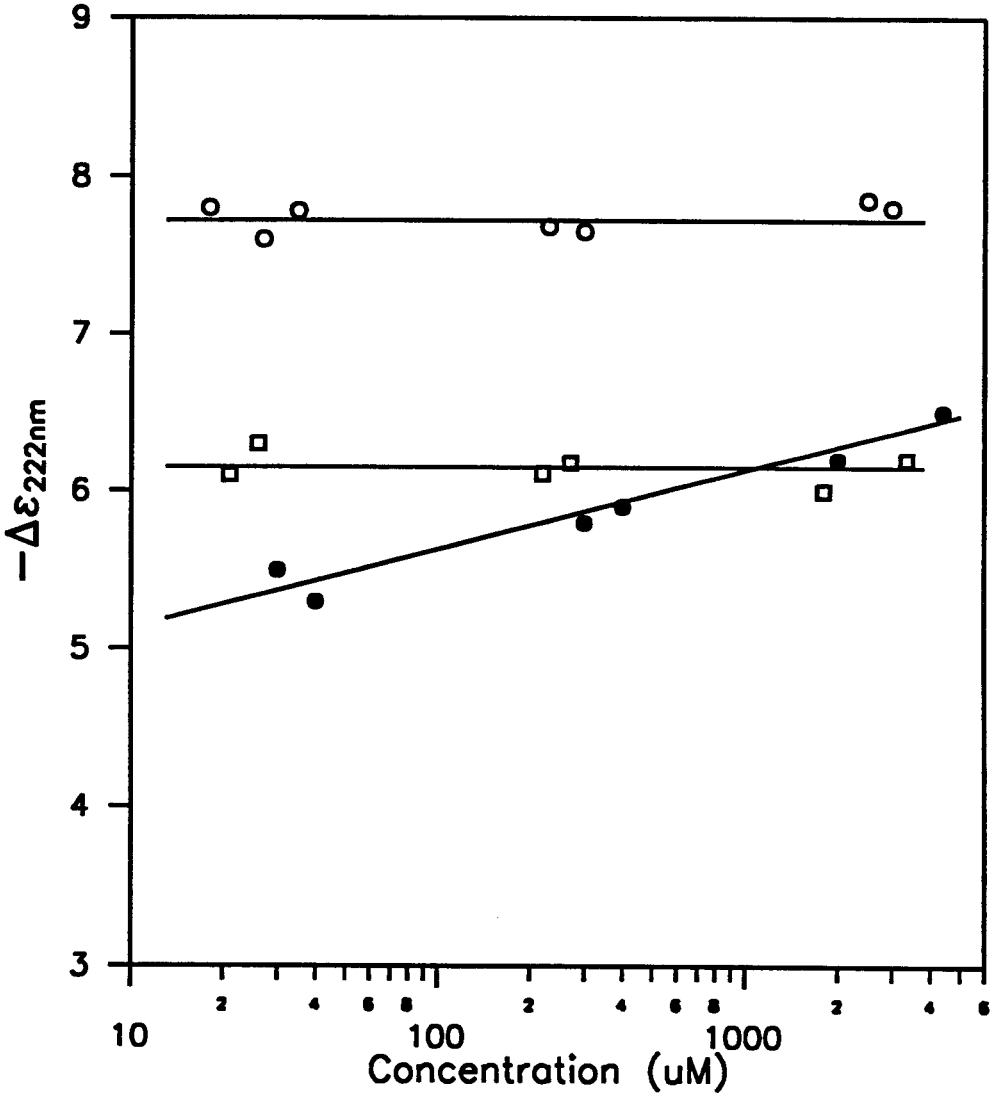
9
8
7
6
5
4
3
$-\Delta\varepsilon_{222nm}$
10
100
1000
Concentration (uM)

Figure 2.5a

Figure 2.5b CD of ERE at 222 nm in TFE as a function of peptide concentration at different temperatures. At 5 °C (-○-), 25 °C (-■-), 45°C (-●-), and 60 °C (-□-).

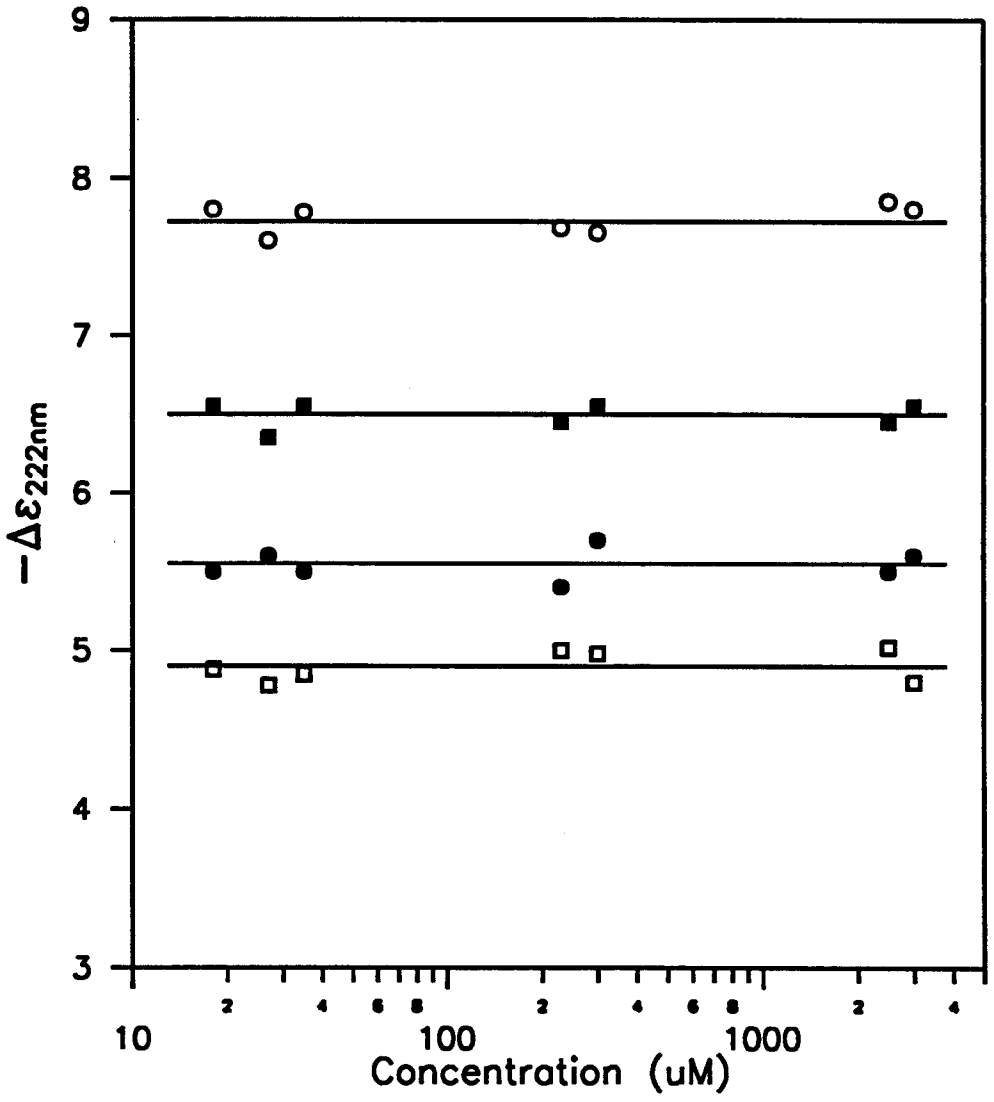
9
8
7
6
5
4
3
$-\Delta\varepsilon_{222nm}$
10
100
1000
Concentration (uM)

Figure 2.5b

Mimicking the Interior Environment of the Protein

Many of the solvents that one might expect to mimic the environment inside a protein caused precipitation of these equivocal peptides, but low percentages of TFE at pH 11, 0.08% digitonin/0.016% cholate with 10 mM phosphate buffer at pH 7.4, 50 mM octylglucoside, and 1 M sucrose in 25 mM MOPS at pH 7.4 gave CD spectra typical of β-strands for some sequences. The hydrophobic environment created by SDS at 2 to 4 mM when the ratio of SDS to peptide is 2:1 to 4:1 consistently gave a large percentage of β-strand.

The characteristic features of a polypeptide in the β-strand conformation are a negative band near 216 nm and a positive band between 195 and 200 nm. However, the CD of a β-strand is much more variable than that of an α-helix, both in amplitude and in the position of the bands. This variability almost certainly results from the much broader range of structures available to β-strands as compared with α-helices (Woody, 1985). Sequences of oligopeptides and proteins can adopt three fundamentally different types of β-sheet conformations: antiparallel, parallel and mixed sheets. The twisting of these different β-form units can perturb the shape of the CD curves (Chothia, 1973). Variation in length and twist of the strands also make the band intensity change (Manning et al., 1988). Manavalan and Johnson pointed out that there is a second set of all-β proteins whose CD spectra actually resemble the spectrum of random coil models (Manavalan and Johnson, 1983). Soybean trypsin inhibitor, wheat germ agglutinin, rubredoxin, elastase, and α-chymotrypsin have a minimum around 200 nm and a positive band in the 185-190 nm range. The crystal structure data of these proteins show that

their β-sheets are either very much distorted or form very short irregular strands (Richardson, 1981). This irregularity may cause the negative CD band to shift from the ideal β-sheet position (210-220 nm) towards the 200 nm region. The presence of this second β-sheet conformation can also complicate CD analysis for secondary structure.

Synthetic polypeptides have been used as models to study β-sheet CD as a pure component. The overall shape of the observed CD curves for poly(Lys), poly(Ser) and poly(Cys) were found to be similar with slight variation. However, due to the significant variation in their amplitudes, these CD curves cannot be regarded as definitive models for β-sheet conformation. More recently oligopeptides such as Boc-(amino-acid)$_n$-OMe with Ala, Val, Leu, and Ile (n usually is larger than 6), were studied as candidates yielding predominantly the β-sheet conformation. Again, although the band positions remain approximately at the same wavelengths, their amplitudes change markedly.

CD measurements on isolated β-sheet sandwich structures have not been accomplished because of the difficulties in synthesizing a well-defined multilayer structure (Manning, 1989). The exact structure of some prepared models is yet to be determined (Moser et al., 1983, Kullman, 1984, Richardson and Richardson, 1987, Unson et al., 1984). Some proteins comprised primarily of a β-sheet sandwich (Chothia, 1984, Chothia and Janin, 1981, Chothia et al., 1981, Chothia and Janin, 1982, Chothia and Lesk, 1982, Chothia, 1983) display widely variant CD spectra. The discrepancies in β-sheet CD are mainly due to variations in strand length, the number of strands

per sheet, and differences in twist and deformation in strands. All of these parameters have been calculated to have a marked effect on the CD spectra of β-sheets (Manning and Woody, 1987, Manning et al., 1988). Theoretical studies indicate that the inter-sheet interactions can modulate CD intensity by up to 10% (Manning et al., 1988). However, the twisting of individual strands and the overall deformation of β-sheets lead to even greater effects on the CD spectra than the intersheet contributions (Manning and Woody, 1987, Manning et al., 1988).

This variable structure is reflected in our β-strand spectra (Figure 2.4a,b,c). The CD of ERE has a negative band around 217 nm ($\Delta\varepsilon$=–3), a positive band around 200 nm ($\Delta\varepsilon$=6.5), and a crossover at 210 nm. Its shape is similar to the CD of the β-strand of poly(Lys), although its positive band is red-shifted from 195 nm to 200 nm. The CD of ADH p15 in the β-strand has a negative band at 224 nm and a positive band at 200 nm which is a red shift of the usual β-sheet CD. The profile resembles that of the β-strands of poly(Val) in water, and suggests the presence of a β-strand although the magnitude deviates from the standard. The CD of CMT p14 as a β-sheet has a negative band at 213 nm and a positive band at 194 nm which is a blue shift of the usual β-sheet CD.

Analysis of the CD (Table 2.8) for ERE (25 °C) gives 60% β-strand, 39% other, and almost no percentage α-helix or β-turn. Analysis for CMT (45 °C) gives 51% β-strand, 38% α-helix, 7% β-turn and 7% other. Analysis for ADH (45 °C) gives about 55% β-strand, 16% α-helix, 3% other and almost no β-turn. The β-strand content in this analysis is lower than that found in the

proteins (86%, 73%, and 56% β-strand). However, our method of analysis is better suited to determining secondary structure in proteins than analyzing a single secondary structure in a short peptide. Furthermore, the CD of a β-strand is much more variable than that of α-helix, as discussed above.

Although the position and amplitude of the bands are variable in the CD of a β-strand (Woody, 1985), the overall CD magnitude from minimum to maximum is fairly constant (15 to 16.5 $\Delta\varepsilon$ units). We can compare the overall magnitude of our β-strand CD to that of poly(Lys) (Greenfield and Fasman, 1969), as an alternate estimate of the percent β-structure. We obtain 62% for ERE, 97% for CMT, and 76% for ADH (Table 2.9).

SDS is a surfactant that can provide a hydrophobic environment for polypeptides in proteins. At high concentrations it forms micelles, and it is well documented that these conditions usually stabilize α-helical structure (Jirgensons, 1977, 1981; Wu and Yang, 1981; Wu et al., 1981; Wu and Yang, 1988; Gierasch, 1989). Yang and coworkers (Wu and Yang, 1981; Wu et al., 1981) successfully used a low concentration (2-4 mM) of SDS to induce β-structure. We followed their methods and dissolved our peptides in aqueous solution without salt, the solutions being self-buffering because of the high concentration of peptide. In the absence of salt the low SDS concentration is far below the critical micelle concentration, CMC (Tanford, 1980). Our equivocal peptides assume the expected α-helical structure in the hydrophilic solvent TFE. The interior of a protein is usually hydrophobic, so we would expect that such an environment would create a β-strand structure for our equivocal peptides, as is found from inspection of the X-ray diffraction for the

Table 2.8 Analysis for secondary structure of the CD of equivocal peptides in 2-4 mM SDS.

Peptide	H	A	P	A+P	T	O	Total
ERE p13	0.00±0.00	0.16±0.05	0.44±0.02	0.60	0.01±0.01	0.39±0.04	1.00
CMT p14	0.38±0.03	0.30±0.04	0.19±0.02	0.51	0.07±0.03	0.07±0.04	1.03
ADH p15	0.01±0.02	0.20±0.03	0.21±0.02	0.41	0.00±0.00	0.57±0.02	0.98

Table 2.9 Comparison of β-strand percentage found in proteins and by different analysis methods.

Peptides	Crystal Structure	SVD/VS	poly(Lys)
ERE p13	0.86	0.60	0.62
CMT p14	0.73	0.51	0.97
ADH p15	0.56	0.54	0.76

parent proteins. The nature of this SDS solvent system is unknown, but we do have 2-4 SDS molecules for each amino acid in the sequence. We presume that the hydrophobic tail of the SDS interacts with the equivocal peptide to mimic the hydrophobic environment found in the interior of the parent protein, while the hydrophilic end of the SDS molecule keeps the β-strand in solution. Here we cannot use CD to test for aggregation because the structure depends on SDS concentration and the ratio of SDS to peptide. We tried to use gradient SDS gel electrophoresis to solve this problem since it can give good resolution to separate small peptides. However, the conditions used here (4 mM SDS, 1 mM sample, 10 mM buffer) are quite different from the conditions used in a normal gel, and we were not able to obtain clear and sharp bands of these peptides with gel electrophoresis.

There is no general agreement about the shape of protein-surfactant complexes; the proposed models can be classified into two major types: Type A, protein is covered with SDS; and type B, protein interacts with SDS micelles. Type A includes these models: 1) Proteins somehow organize SDS into micellar complexes of definite size and shape. This model was based on results of SDS gel electrophoresis (Maizel, 1969). 2) The 'rodlike -particle model' was proposed by Reynolds and Tanford (1970) based on hydrodynamic measurements. 3) A flexible 'α-helix/random coil' with random coil properties was suggested by Mattice et al. (1976). They used CD to measure protein binding with SDS, and found that α-helix content increases while β-strand decreases. These results are also confirmed by synthetic polypeptides having Lys, Arg and His residues. Type B includes: 1) The

'necklace model': the polymer chain is flexible, and micelle-like clusters of SDS are scattered along the chain. This was derived from free boundary electrophoresis experiments in which SDS-protein complexes migrated at about the same velocity as SDS micelles (Shirahama et al., 1974, Takagi et al., 1975). In the improved version of this model, SDS binds to the protein in the form of spherical micelles, causing it to form α-helices mostly in the hydrophobic region of the micelles (Rao, 1989). 2) A model which winds Mattice's α-helix/random structure around single cylindrical micelles is based on no new experiments (Lundahl et al., 1986) 3) Protein-decorated micelle structures of SDS that explain neutron scattering data (Ibel et al., 1990). It will be interesting to further investigate the structure of protein-SDS complexes since there are still many unanswered questions. However, some facts can be derived from these studies, such as: 1) SDS destroys native structure but supports secondary structure, primarily α-helix. 2) Short peptides behave differently, and may have different secondary structures at different SDS concentrations (below or above the CMC).

Since SDS can stabilize a β-strand in these equivocal peptides, we thought that octanol might be a good candidate as a solvent system to stabilize β-strand; it resembles the structure of SDS with a long aliphatic group. It is also capable of hydrogen bonding to any exposed peptide functional groups, yet is quite hydrophobic. However, our peptides did not dissolve in octanol. Simply partitioning the peptide between octanol and a water phase failed, because octanol and water are immiscible. Ostermen and Kaiser (1985) reported that an amphiphilic synthetic peptide TFA β-13 can

dissolve in octanol with a small aliquot of 0.1 N HCl (aq) added, and can form a β-strand in octanol as it does in TFE and buffer. Under these conditions, our equivocal peptides did not dissolve, and we were not able to use octanol to stabilize a β-strand for the equivocal peptides.

When ERE p13 is in 20% TFE, it is an α-helix as judged by CD. In pH 9.0 water it starts to show a transition to β-strand. In 16% TFE, 84% water at pH 11, its CD is that of a β-strand (Figure 2.6). At an amide concentration of 3 mM, its CD is more like the conventional CD of a β-strand as shown by poly(Lys), with a positive band at 195 nm instead of 200 nm as in SDS, and a negative band at 216 nm. Using the overall magnitude method for comparison, it contains 75% β-strand. However at a lower concentration, about 1 mM amide, its CD shows only some β-strand features, with a negative band at about 220 nm ($\Delta\varepsilon$=-3.5), and a positive band about 205 nm ($\Delta\varepsilon$=1). In addition, there is another positive band at about 190 nm ($\Delta\varepsilon$=1.1). It is rather similar to the CD of TNF-α measured recently in our laboratory (220 nm $\Delta\varepsilon$=-1 , 203 nm $\Delta\varepsilon$=1.6 and 192 nm $\Delta\varepsilon$=1.1) (Riazance-Lawrence *et al.*, 1991), which is a β-protein with 50% β-strand and no α-helix. At 0.1 mM amide the intensity of the CD increases, and the position of the bands shift a little, with a positive band at 195 nm and a negative band at 218 nm. The intensity of both positive and negative bands is larger than for a typical β-strand CD. Since the CD spectra change with concentration, this demonstrates that the β-strand observed under these conditions is intermolecular. At 1 mM, the solution is turbid, indicating some aggregation at this concentration; however,

Figure 2.6 CD of ERE as a function of concentration in 16% TFE, 84% water at pH 10. ERE amide concentration is : (1) 3 mM (—), (2) 1 mM (---), and (3) 0.1 mM (...).

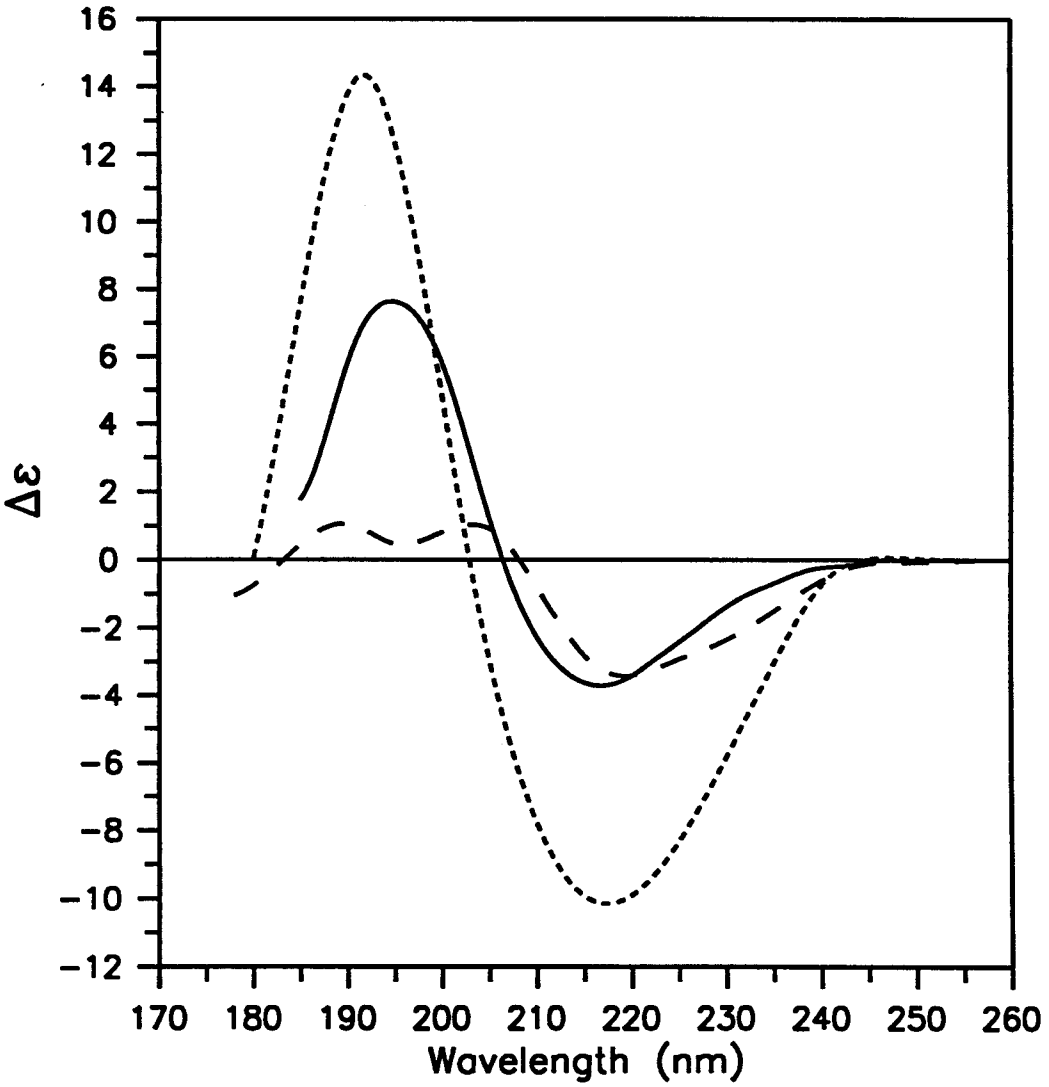
16
14
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
−12
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.6

when the concentration increases or decreases, the solution is not turbid. In fact, in its crystal structure the peptide forms an antiparallel β-strand with another sequence. For CMT p14 and ADH p15, under the same conditions, the α-helix content decreases, but an α-helix-like CD is still retained. Other researchers have also shown that proteins or peptides can form β-strands in low percentages of TFE. At about 30% TFE and intermediate pH (4 to 7), naturally occurring synthetic amyloid β-peptides can form β-structures (Barrow and Zagorski, 1991).

We analyzed the hydrophobicities of the three equivocal amino acid sequences using the scale developed by Kyte and Doolittle (1982). The helical wheel representation of the α-helix and the β-strand for the sequences do not show amphipathic characteristics. Perhaps this non-amphipathic property also makes these sequences more "equivocal", instead of becoming normal α-helices in the parent proteins.

Solvent Effects on a Random Structure Sequence

Since equivocal amino acid sequences can form an α-helix or β-strand in defined solvent systems, we would like to ask the following question: will TFE promote an α-helix, or SDS a β-strand for any sequence? In order to answer this question we chose a sequence with a preference for irregular secondary structure. Sequence HIT 2-23 is from HIV-1 Tat protein. Tat protein is a trans-acting transcriptional activator of human immunodeficiency virus type 1, and is essential for viral transcription. The Chou and Fasman method showsthat this HIT sequence has neither α-helix nor β-strand forming

potential, therefore it is a good candidate for our study. It contains a potential protein activating region, residues 2-11 (Rappaport et al., 1989). Loret et al. (1991) showed that this sequence in buffer or 90% TFE gives a random CD spectrum. The two spectra show a negative band at 197 nm typical of random coil and a small negative band at 220 nm, which could be due to the prolines, as in random-coil collagen (Jenness et al., 1976), or to the aromatic ring of the tryptophan (Woody, 1985). They called the spectrum in pH 7 buffer, type A and the one in 90% TFE type B. Type A has a larger negative band at 197 nm and a smaller negative band at 220 nm (Figure 2.7).

The CD spectrum of peptide HIT was also studied in different hydrophobic solvents, in aqueous solution at different pHs and at different temperatures (Loret et al., 1991). A mixture of 70% ethanol and 30% water, as well as a mixture of 30% ethanol, 30% cyclohexane, and 40% octanol gives a random-like spectrum similar to the type B CD. At 20 °C, the spectrum at pH 3 and 9 is a random-like spectrum similar to a type A spectrum in Figure 2.8. Interestingly, temperature has different effects at different pHs. From 2° to 65 °C, the CD spectrum at pH 3 and 9 remains unchanged (type A). In contrast, the CD at pH 7 is a type B-like spectrum above 45 °C. This change occurs between 37° and 45 °C. These experiments indicate that in aqueous solution the uncharged amino acids block the temperature effect. The spectrum in 70% ethanol, pH 9 and 45 °C is similar to the one in 90% TFE (type B). There are five prolines in this sequence, and their slow trans-cis isomerization might influence the secondary structure. However, the CD spectrum of peptide 2-23 was also measured

after different incubation times (2, 6, 24, and 48 hours), and the results show that spectra obtained were independent of the time.

We placed sequence HIV 2-23 into the solvent system that stabilizes the β-strand for our equivocal peptides, 3 mM SDS. The CD spectrum remains random-like, with a slightly larger intensity around the shoulder at 220 nm and at 185 nm, and a negative band of almost the same intensity as type A at 200 nm. The CD data were analyzed for the percentage of secondary structure (Table 2.9). This analysis gives a typical random coil characteristic for all three different systems. This is a baseline experiment showing that TFE and SDS do not necessarily promote secondary structure in a sequence regardless of its amino acid preference.

Figure 2.7 Random-like CD of HIT in different solvent systems: (1) 20 mM phosphate buffer, pH 7.0 (——), (2) 90% TFE (---), and (3) 3 mM SDS (...).

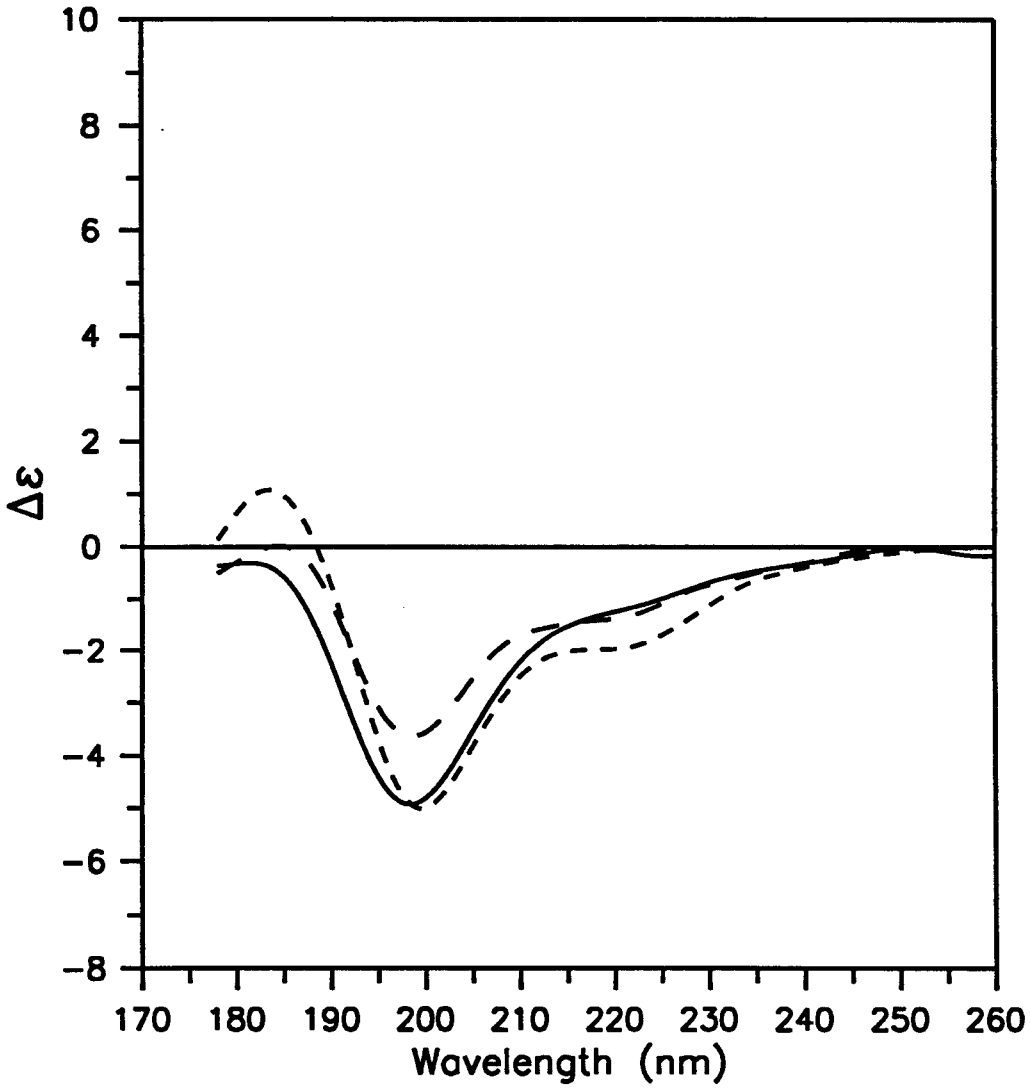
10
8
6
4
2
0
−2
−4
−6
−8
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2.7

Table 2.10 Secondary structure analysis of peptide HIT in buffer, 90% TFE, and SDS.

solvent	H	A	P	T	O	total
aqueous	0.13	0.21	-0.02	0.30	0.39	1.03
90% TFE	0.13	0.21	0.00	0.25	0.42	1.00
3 mM SDS	0.16	0.15	-0.02	0.28	0.42	1.01

CONCLUSIONS

Our results are consistent with recent work (Bowie *et al.*, 1991), which showed that the inverse protein folding problem can be effectively attacked by finding sequences that are most compatible with the environments of the residues in their three-dimensional structure. They reduced the 3D structure to a one-dimensional string of environmental classes at the sequence level, and were able to detect a structural similarity between actins and 70-kilodalton heat shock proteins, even though these protein families share no detectable sequence similarity. Here we show that equivocal amino acid sequences can be made to assume both the α-helix or the β-strand conformation by proper choice of the solvent system. The results indicate that there is a common thread in the behavior of these equivocal amino acid sequences, since they all can form a stable α-helix in 100% TFE at 0°C, and a β-strand in low concentrations of SDS.

Our results demonstrate that the solvent is in control; therefore, schemes that predict secondary structure from primary structure alone can never be totally successful. Tertiary structure must be taken into account so that we know what solvent the peptide sequence effectively sees when protein folding is complete. Since the environment can indeed change the secondary structure, the hierarchic model will have to be modified to take folding feedback into account: (1) primary structure determines secondary structure, (2) secondary structures fold into a tertiary structure, (3) some secondary

structures change as a result of their new environment, and (4) minor rearrangements occur in the tertiary structure. On the other hand, these results fit in with Dill's non-hierarchic model (1990), which involves random condensation and then segment rearrangement, and can contribute to folding simulations (Cohen et al., 1986; Skolnick and Kolinski 1990). Other researchers now also attempt to incorporate tertiary structure into their predictions by using concepts such as amphipathicity (Finer-Moore et al. 1989) and hydrophobic moments (Rose and Dworkin, 1989; Eisenberg et al. 1989). If we are going to understand how proteins fold, we must understand the environmental as well as the sequence effects.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM-21479 from the National Institute of General Medical Sciences. It is a pleasure to acknowledge Ping-Jung Chou for his smoothing program that uses the cubic spline algorithm, and Dean Malencik for his help with the HPLC.

REFERENCES

Balccrski, J. S., Pysh, E. S., Bonora, G. M. and Toniolo, C. (1976) J. Am. Chem. Soc. 98, 3470-3474.

Barrow, C. J. and Zagorski, M. G. (1991) Science 253, 179-182.

Bowie, J., Luthy, R. and Eisenberg, D. (1991) Science 253, 164-170.

Brandts, J. F. and Hunt, L. (1967) J. Am. Chem. Soc. 89, 4826-4838.

Bruch, M. D., McKnight, C. J. and Gierasch, L. M. (1989) Biochemistry 28, 8554-8561.

Burgess, A. W., Ponnuswamy, P. K. and Scheraga, H. A. (1974) Isral J. Chem. 12, 239-286.

Chen, G. C. and Yang, J. T. (1977) Anal. Lett. 10, 1195-1206.

Chothia, C., Levitt, M., and Richardson, D. (1981) J. Mol. Biol. 145, 215-250.

Chothia, C. and Janin, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4146-4150.

Chothia, C. and Janin, J. (1982) Biochemistry 21, 3955-3965.

Chothia, C. and Lesk, A. M. (1982) J. Mol. Biol. 160, 309-323.

Chothia, C. (1983) J. Mol. Biol. 163, 107-117.

Chothia, C. (1984) A Rev. Biochem. 53, 537-572.

Chou, P. Y. and Fasman, G. D. (1974a) Biochemistry 13, 211-222.

Chou, P. Y. and Fasman, G. D. (1974b) Biochemistry 13, 222-245.

Chou, P. Y. and Fasman, G. D. (1978a) Adv. Enzymol. 47, 45-148.

Chou, P. Y. and Fasman, G. D. (1978b) Ann. Rev. Biochem. 47, 251-276.

Dill, K. A. (1990) Biochemistry 29, 7133-7155.

Edelman, J. and White, S. H. (1989) J. Mol. Biol. 210, 195-209.

Eisenberg D. Wesson M. and Wilcox W. (1989) Predictions of Protein Structure and the Principles of Protein Conformation, (Fasman, G. D. ed.) Plenum Press, New York. pp. 635-646.

Elwell, M. L., Schelleman, J. A. (1977) Biochim. Biophys. Acta 494, 367-383.

Epand, R. M. and Scheraga, H. A. (1968) Biopolymers 6, 1551-1571.

Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.

Gibrat, J.-F., Garnier, J. and Robson, B. (1987) J. Mol. Biol. 198, 425-443.

Gierasch, L. M. (1989) Biochemistry 28, 923-930.

Goodman, M., Verdini, A. S., Choi, N.S. and Masuda, Y. (1970) in Topics in Stereochemistry (Eliel, E. and Allinger, N.L. ed.), John Wiley, New York, pp. 69-166.

Helenius, A. and Simons, K. (1975) Biochim. Biophys. Acta 415, 29-79.

Hennessey, J. P. Jr. and Johnson, W. C. Jr. (1981) Biochemistry 20, 1085-1094.

Hennessey, J. P. Jr., Johnson, W. C. Jr., Bahler, C. and Wood, H. G. (1982) Biochemistry 21, 642-646.

Hennessey, J. P. Jr., Manavalan, P., Johnson, W. C. Jr., Malencik, D. A., Anderson, S. R., Schimerlik, M. I. and Shalitin, Y. (1987) Biopolymers 26, 561-571.

Ho, S. P. and Degrado, W. F. (1987) J. Am. Chem. Soc. 109, 6751-

Holladay, L. A. and Wilder, P. (1980) Biochim. Biophys. Acta 629, 156-167.

Holley, L. H. and Karplus, M. (1989) Proc. Natl. Acad. Sci. USA 86, 152-156.

Ibel, K., May, R. P., Kirschner, K., Szadkowski, H., Mascher, E. and Lundahl, P. (1990) Eur. J. Biochem. 190, 311-318.

Jirgensons, B. (1977) Biochim. et Biophys 473, 352-358.

Jirgensons, B. (1981) Makromol. Chem. Rapid Commun. 2, 213-217.

Johnson, W. C. Jr. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 145-166.

Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577-2637.

Kelly, M. M., Pysh, E. S., Bonora, G. M. and Toniolo, C. (1976) J. Am. Chem. Soc. 99, 3264-3266.

Knecht, R. and Chang, J. Y. (1986) Anal. Chem. 58, 2375-2379.

Kullman, W. (1984) J. Med. Chem. 27, 106-115.

Lau, S. Y. M., Taneja, A. K. and Hodges R. S. (1984) J. Chromatogr. 317, 129-140.

Lehrman, S. R., Tuls, J. L. and Lund, M. (1990) Biochem. 29, 5590-5596.

Levin, J. M., Robson, B. and Garnier, J. (1986) FEBS Lett. 205, 303-308.

Lim, V. I. (1974a) J. Mol. Biol. 88, 857-872.

Lim, V. I. (1974b) J. Mol. Biol. 88, 873-894.

Lundahl, P., Greijer, E., Sandberg, M., Cardell, S. and Erikssion, K. -O. (1986) Biochim. Biophys. Acta 873, 20-26.

Maizel, J. V. Jr. (1969) in Fundamental Techniques in Virology (Habel, K. and Salzman, N. P., ed.) pp. 334-362, Academic Press, New York.

Malencik, D. A., Zhao, Z., and Anderson, S. R. (1990) Anal. Biochem. 184, 353-359.

Manavalan, P., Johnson, W. C. Jr. and Modrich, P. (1984) J. Biol. Chem. 259, 11666-11667.

Manavalan, P. and Johnson, W. C. Jr. (1985) Biochim. Biophys. Acta 829, 365-370.

Manavalan, P., Mittelstaedt, D. M., Schimerlilk, M. I. and Johnson, W. C. Jr. (1986) Biochemistry 25, 6650-6655.

Manavalan, P. and Johnson, W. C. Jr. (1987) Anal. Biochem. 167, 76-85.

Manning, M. C. and Woody, R. W. (1987) Biopolymers 26, 1731-1752.

Manning, M. C., Illangeskare, M. and Woody, R. W. (1988) Biophys. Chem. 31, 77-86.

Manning, M. C. (1989) J. of Pharmaceutical and Biomedical Analysis 7, 10, 1103-1119.

Mattice, W. L., Riser, J. M. and Clark, D. S. (1976) Biochemistry 15, 4264-4272.

McClarin, J. A., Frederick, C. A., Wang, B-C., Green, P., Boyer, H. W., Grable, J. and Rosenberg, J. M. (1986) Science 234, 1526-1541.

Merutka, G. and Stellwagen, E. (1989) Biochemistry 28, 352-357.

Moser, R., Thomas, R. M. and Gutte, B. (1983) FEBS Lett. 157, 247-251.

Mutter, M. and Hersperger, R. (1990) Angew. Chemie Int. Ed. Eng. 29, 185-187, 1990.

Narayanan, U., Keiderling, T. A., Bonora, G. M. and Toniolo, C. (1986) J. Am. Chem. Soc. 108, 2431-2437.

Nelson, J. W. and Kellenbach, N. R. (1986) Proteins: Struct., Funct., Genet. 1, 211-217.

Nelson, J. W. and Kallenbach, N. R. (1989) Biochemistry 28, 5256-5261.

Nishikawa K. and Ooi, T. (1986) Biochim. Biophys. Acta 871, 45-54.

Osternman, D. G. and Kaiser, E. T. (1985) J. of Cellular Biochem. 29, 57-72.

Palau, J., Argos, P. and Puigdomenech, P. (1982) Int. J. Peptide Protein Res. 19, 394-401.

Pongor, S. and Szaley, A. A. (1985) Proc. Natl. Acad. Sci. USA 82, 366-370.

Qian, N. and Sejnowski, T. J. (1988) J. Mol. Biol. 160, 865-884.

Rao, P.-F. (1989) Doctoral Thesis, Faculty of Science, Osaka University.

Reynolds, J. A. and Tanford, C. (1970) Proc. Natl. Acad. Sci. USA 66,

1002-1007.

Richardson, J. S. and Richardson, D. C. (1987) in Protein Engineering (D. L. Oxender and C. F. Fox, eds.), pp.149-163, Liss, New York.

Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327-356.

Rose, G. D. and Dworkin, J. E. (1989) Predictions of Protein Structure and the Principles of Protein Conformation, (Fasman, G. D. ed.) Plenum Press, New York. pp. 625-634.

Rosenheck, K. and Doty, P. (1961) Biochemistry 47, 1775-1785.

Shirahama, K., Tsujii, K. and Takagi, T. (1974) J. Biochem. (Tokyo) 75, 309-319.

Skolnick, J. and Kolinski, A. (1990) Science 250, 1121-1125.

Sweet, R.M. (1986) Biopolymers 25, 1565-1577.

Takagi, T., Tsujii, K. and Shirahama, K. (1975) J. Biochem. (Tokyo) 77, 939-947.

Tanford, C. (1968) Adv. Protein Chem. 23, 121-282.

Tanford, C. (1980) "The Hydrophobic Effect: Formation of Micelles and Biological Membranes" Wiley-Interscience, p.66-68.

Unson, C. B., Erickson, B. W., Richardson, D. C. and Richardson, J. S. (1984) Fed. Proc., Fed. Am. Soc. Exp. Biol. 43, 1837.

von Hippel, P. H. and Wong, K. -Y. (1965) J. Biol. Chem. 240, 3909-3923.

von Hippel, P. H. and Schleich, T., The Effects of Neutral Salts on the Structure and Conformational Stability of Macromolecules in Solution., Structure and Stability of Biological Macromolecules, Timasheff, S. N., and Fasman, G. D., eds. Dekker Inc., N.Y. (1969).

Weber, K. and Osborn, M. (1969) J. Biol. Chem. 244, 4406-4412.

Woody, R. W. (1985) The Peptides 7, 15-114.

Wu, C. C. and Yang, J. T. (1981) Mol. Cell. Biochem. 40, 109-122.

Wu, C. C., Ikeda, K. and Yang, J. T. (1981) Biochemistry 20, 566-570.

Wu, C. C. and Yang, J. T. (1988) Biopolymers 27, 423-430.

Zvelebil, M. J., Barton, G. J., Taylor, W.R. and Sternberg, M. J. E. (1987) J. Mol. Biol. 195, 957-961.

SECTION III

Poly[d(Gm^{5}C).d(Gm^{5}C)] CAN ASSUME THE Z' FORM: A CD STUDY

Lingxiu Zhong and W. Curtis Johnson, Jr.

Department of Biochemistry and Biophysics
Oregon State University
Corvallis, OR 97331-6503

Published in Biopolymers (1990), 30, 821-828.

ABSTRACT

Poly[d(Gm^{5}C).d(Gm^{5}C)] can be titrated from the Z-form in 30% ethanol to the Z'-form by adding either ethanol or divalent cations. Analysis by singular value decomposition of CD spectra recorded during the titration reveal that ethanol and transition metal induced changes are two state with a single step titration curve. When the change is induced by alkaline earth metals it is still two state, but in contrast the titration curve is complex indicating two binding sites.

INTRODUCTION

Since the discovery of Z-form DNA in crystals (Wang et al., 1979) and in solution (Pohl and Jovin,1972), research has shown that there is not just one Z-DNA but rather a family, Z-, Z_I-, Z_{II}-, and Z'-DNA, the particular form depending on environmental conditions. Two main forms, Z_I and Z_{II}, are observed in the crystal structure of the hexanucleotide d(GpC), depending essentially on the coordination of the phosphate group to water or to a hydrated magnesium ion (Wang et al., 1981, Gessner et al., 1989). In the Z'-DNA tetramer d(CGCG), the deoxyguanosine sugar is puckered at C_1-exo rather than at C_3'-endo found in Z-, Z_I-, and Z_{II}-DNA (Drew et al., 1980).

A second solution form of poly[d(GC).d(GC)] was described by Pohl (Pohl, 1976), and later by Hall and Maestre (Hall and Maestre, 1984), who also called this the Z'-form. They discovered that it is produced at a high ethanol concentration (85% v/v), and the Z to Z' transition is a function of alcohol. Recently, it has been shown that the Z'-form of left-handed poly[d(GC).d(GC)] in solution depends on the presence of multivalents, and thus can be associated with the Z_{II}-form in crystals (Harder and Johnson, 1990).

The methylation of DNA is known to play an important role in gene regulation (Razin and Riggs, 1980, Doerfler, 1983), and poly[d(Gm^5C).d(Gm^5C)] can undergo a B to Z transition under physiological conditions (Behe and Felsenfeld, 1981). It has remarkable conformation lability, which is extremely sensitive to the presence of trace metal ions, salt

concentration, temperature, or ligand binding (Chen, 1986, Chaires, 1985, Devarajan and Shafer, 1986).

Here we study the Z to Z' transition of poly[d(Gm^{5}C).d(Gm^{5}C)] in solution during alcohol and divalent ion titration. We find that ethanol and transition metals can promote the Z'-form of this polymer as they do for poly[d(GC).d(GC)] (Harder and Johnson, 1990), but in contrast, ethanol alone is sufficient to promote the Z'-form. Furthermore, binding curves for alkaline earth metals show two binding sites that promote the Z'-form of this polymer.

MATERIALS AND METHODS

Sample Preparation

Disodium ethylenediaminetetraacetic acid (99.5% pure from Sigma) and tetrasodium ethylenediaminetetraacetic acid (99% pure from Sigma) were mixed together to make trisodium ethylenediaminetetraacetic acid (Na_3EDTA), some at pH 8.0 and some at pH 7.0. MOPS (3-morpholino propanesulfonic acid) as the free acid and sodium salt (gold label, Research Organics Inc.) were mixed as MOPS buffer, some at pH 7.0 and some at pH 8.0. Thus MOPS and EDTA buffers were prepared without using sodium hydroxide, which would change the sodium ion concentration and often contains impurities. Sodium chloride was obtained from Fluka, 99.5% pure. Ethanol (USI Chemicals Co.) was passed over a Chelex column before use to eliminate all divalent ions, then stored in teflon bottles (treated ethanol was not used unless it had an absorbance at 230 nm of less than 0.03/cm). All water used originated from a Milli-Q reagent water system (Millipore Corp.), had a resistivity of 18 megaohm-cm and was dithizone-negative, indicating it was free of metals. All labware used was plastic. $CaCl_2$ (Baker Chemical Co.), $MgCl_2$ (99.3% pure from Mallinckrodt), $CoCl_2$ (98% pure from Aldrich), $NiCl_2$ (99% pure from Aldrich) and $ZnCl_2$ (98% pure from EM Science) were used without further purification. Divalent ion stock solutions were from 1 to 5 mM for titration purposes.

Poly[d(Gm^5C).d(Gm^5C)] was purchased from Pharmacia (5 units) and dissolved in 1 ml H_2O. The polynucleotide was dialyzed using Spectra/por

tubing (6000-8000 molecular weight cut-off) against 1 liter 0.5 M NaCl, 10 mM EDTA, pH 8.0 for 24 hours to remove divalent impurities. This was followed by dialysis into 1 liter of 1 mM MOPS, 0.1 mM EDTA, pH 8.0. To make sure that all divalent ions were removed from the DNA we used two columns. AG50w-x8 cation exchange resin (Bio-Rad Chemical Division) was packed in a 2 ml column, washed with 1 liter of distilled and deionized water, and regenerated by 2 liters of 0.05 N HCl. The column was converted to Na^+ with 1 liter of 100 mM NaCl followed by 1 mM MOPS, 0.1 mM EDTA, pH 7.0. Also, Chelex-100 (Bio-Rad Laboratories) was packed in a 10 ml column, and washed with 1 liter of 1 mM MOPS, 0.1 mM EDTA, pH 7.0 buffer. The stock DNA was then passed through either one of these columns. The DNA was concentrated in a Centricon microconcentrator (centrifuged at 5000 rpm for 3 hours). 2 ml of 1 mM MOPS, 0.1 mM EDTA, pH 7.0 was added to the sample, then reconcentrated as above. Finally the DNA was recovered in the Centricon by centrifuging at 2000 rpm for 2 mins, and its volume was adjusted to about 100 ml. DNA concentration, always given on a per nucleotide basis, was determined spectrophotometrically using an ε(256 nm) of 7100 (Devarajan and Shafer, 1986). The stock solution was stored on ice at 4°C.

With these dialysis and ionic exchange procedures divalent ions should be almost completely removed. Mg^{2+} and Ca^{2+} are expected to be less than 2 and 5 ions per 1000 nucleotides, respectively (Devarajan and Shafer, 1986). The CD spectrum shows that DNA in 1 mM MOPS, pH 7.0 is indeed in the B-form, which agrees with the data published by other investigators (Behe and Felsenfeld, 1981).

Circular Dichroism

CD spectra in the range 320-210 nm were collected every 0.5 nm on a Jasco-J40 spectrophotometer interfaced with a Leading Edge PC computer. From 320 nm to 240 nm, we collected the voltage at each point corresponding to a 4-second time constant, while from 239.5 nm to 210 nm we used the equivalent of a 16-second time constant. Measurements were partially repeated and extended from 220 nm to 184 nm using a vacuum ultraviolet (VUV) CD spectrophotometer (Johnson, 1971) and an automated data collection system consisting of a Leading Edge model D computer and a Metrabyte Dascon-1 a/d board. The signal to noise ratio was controlled by changing the scanning speed of the monochromator, which in these experiment was 0.5 nm/min. Both instruments were calibrated with (+)-10-camphorsulfonic acid (CSA) using a $\Delta\varepsilon$(290.5 nm) of +2.37 $M^{-1}cm^{-1}$ (Chen and Yang, 1977). Software for both instruments were written in this laboratory. UV spectra were recorded on a Cary 15 spectrophotometer. All spectra were recorded at 20°C.

For the ethanol and divalent ion titration experiments most DNA concentrations were 30 mM - 50 mM (nucleotide basis) in a 1 cm pathlength cell. For ethanol at 85%, DNA concentrations were about 20 mM (OD = 0.15) to avoid precipitation. None of the absorbance spectra showed any apparent absorption at wavelengths outside of the absorbance bands, indicating that no aggregation occurred. For the VUV CD measurements a 100 μm cell was used and the DNA concentration was about 100 mM.

Ethanol was added dropwise to the samples as they were stirred

rapidly. The ethanol concentration was incremented from 10% to 30% (v/v) and CD spectra were recorded after 45 mins. The spectra show that the Z-form transition is not complete until the solution contains 30% (v/v) ethanol. When the percentage of ethanol is further increased, poly[d(Gm^5C).d(Gm^5C)] starts to undergo a Z to Z' transition. In this way we produced the Z-form of poly[d(Gm^5C).d(Gm^5C)] by adding ethanol to DNA in 1 mM MOPS, pH 7.0 to a final concentration of 30% v/v. This mixture was allowed to stand for approximately 1/2 hour to allow the B to Z transition to equilibrate. Ethanol concentration is then increased or metal chlorides are added for a titration to the Z'-form.

CD Data Analysis

Singular Value Decomposition Theory. The CD spectra from ethanol and divalent ion titrations were analyzed by using a singular value decomposition (SVD) method developed in our laboratory. SVD is particularly useful for finding the number of independent components in a set of spectra, and for averaging the set of spectra to remove both high frequency noise and low frequency random error. Basic theory (Noble and Daniel, 1977, Lloyd, 1969) and an application of this method for CD titration spectra are discussed elsewhere (Harder and Johnson, 1990, Johnson, 1985). Briefly, spectra sampled at m wavelengths are placed in a rectangular matrix, A, of n columns, one column for each spectrum collected in a titration. This matrix is decomposed into the product of three matrices $\mathbf{A} = \mathbf{USV}^T$: **U**, a unitary column matrix of orthogonal basis vectors; **S**, a matrix with singular values on the

main diagonal; $\mathbf{V}^T$, a unitary row matrix of weighting coefficients whose entries fit the eigenvector matrix to the original data set. The singular values are the positive square roots of the eigenvalues common to both $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$. We weigh each eigenvector in **U** with its corresponding singular value to form a column matrix, **US**, of basis CD spectra that are orthogonal.

Information Content. The singular values in the **S** matrix weight the importance of each orthogonal basis vector in **U**. The number of independent components in a set of data is the number of singular values which are significant above noise. Real data are noisy, and there is usually no singular value that is identically zero. For instance, in our Ca^{2+} titration spectra, two significant singular values will give an information content of two, which means that there are only two independent species with CD spectra in matrix **A** that are significant above the noise. Therefore, this is a two-component system, although the set of data does not show an isosbestic point. Here the **S** matrix obtained from SVD is important to us since it gives the number of independent variables above the noise in a series of spectra.

Averaging Among Spectra in a Series. SVD allows averaging among different spectra in a series, so that it is not necessary to repeat each spectrum in a series in order to obtain an average spectrum. Once we obtain the significant singular values, we can determine the corresponding significant basis CD spectra in **US**. The original data can then be reconstructed using only the significant basis CD spectra and the appropriate weighting coefficients in $\mathbf{V}^T$ to yield data that are effectively averaged among the series of spectra. We find the number of basis CD spectra that are significant above the noise

by inspection; this is the number of independent components in the system. We smooth the original data by forming linear combinations of significant basis vectors with the corresponding columns of $\mathbf{V}^T$. The smoothed CD spectra reconstructed in this way average the random errors in measurement as well as the high frequency noise.

SVD averaging is interspectral smoothing, which removes noise at each wavelength uncorrelated among the experimental spectra. Each spectrum can be further smoothed by a weighted averaging which combines data over a range of wavelengths. This is the intraspectral smoothing that is commonly used to remove noise uncorrelated within each spectrum. In order to remove fluctuations that are uncorrelated within each spectrum, our reconstructed CD spectra were further smoothed using a 17-point Savitzky-Golay procedure (Savitzky and Golay, 1964) while VUV CD spectra were smoothed using a Fourier Transform smoothing program (Aubanel and Oldham, 1985). VUV CD spectra were adjusted to fit the near UV CD spectra using a linear least-squares fit of redundant data in the 210 nm - 220 nm region.

Following Variables in a Series of Spectra. The **V** matrix contains the least squares coefficients which fit the basis CD spectra in **US** to the original data. The most significant basis CD spectrum represents the common features in the series; the least important basis CD spectrum that is still significant represents the most changeable features in the series above the noise. Thus the coefficients in **V** for the least important but significant basis CD spectrum monitor the changes that are occurring in the series as

integrated over the entire wavelength range. Furthermore, the changes in experimental spectra because of the changes in titrant concentrations are reflected in the smooth changes in the fitting coefficients found in the columns of **V**, which correspond to the significant basis vectors. Therefore, a plot of the **V** columns versus titrant concentration can give us information about the titration. The **V** matrix is useful to us because it provides data points for melting or titration curves which are integrated over entire spectra rather than fixed at one wavelength. We can fit these curves to models to deduce thermodynamic parameters and also use these data to reconstruct the original spectra to further eliminate experimental errors.

RESULTS AND DISCUSSION

Ethanol Titration

Davarajan and Shafer (1986) have shown that a Z-form CD spectrum for poly[d(Gm^{5}C).d(Gm^{5}C)] in low salt is due to contamination by divalent cations. Our CD spectrum in Figure 3.1 demonstrates that our specially treated poly[d(Gm^{5}C).d(Gm^{5}C)] is free of divalents, as it is in the B-form in low salt, 1 mM MOPS pH 7.0. Figure 3.1 also shows that titrating to 30% (v/v) ethanol converts the polymer into the Z-form (Behe and Felsenfeld, 1981).

When the percentage of ethanol is increased from 30% to 80% (Figure 3.2), the negative band around 295 nm, characteristic of the near-UV CD spectrum of Z DNA, decreases in magnitude. When the percentage is raised to 85%, this band disappears. The spectrum around 250-230 nm also changes during titration from positive to negative; a positive band emerges at 283 nm ($\Delta\varepsilon$ = 4.47) and there is an isosbestic point at 276 nm. These characteristics demonstrate that we have produced the Z'-form, studied in poly[d(GC).d(GC)] by Hall and Maestre (1984). The Z' spectrum for poly[d(GC).d(GC)] is characterized by a positive band at 278 nm ($\Delta\varepsilon$ = 5.56) and the absence of the negative band at 292 nm; the transition also has an isosbestic point at 268 nm.

The isosbestic point indicates that we have a two-state transition, and this is confirmed by SVD analysis of these titrations (Figure 3.3). These basis vectors show that the first two carry the most significant information, since the largest three singular values are 83.05, 48.46, and 9.57. Furthermore, the

Figure 3.1 CD spectra of the B- to Z-transition with ethanol for poly[d(Gm^{5}C).d(Gm^{5}C)].

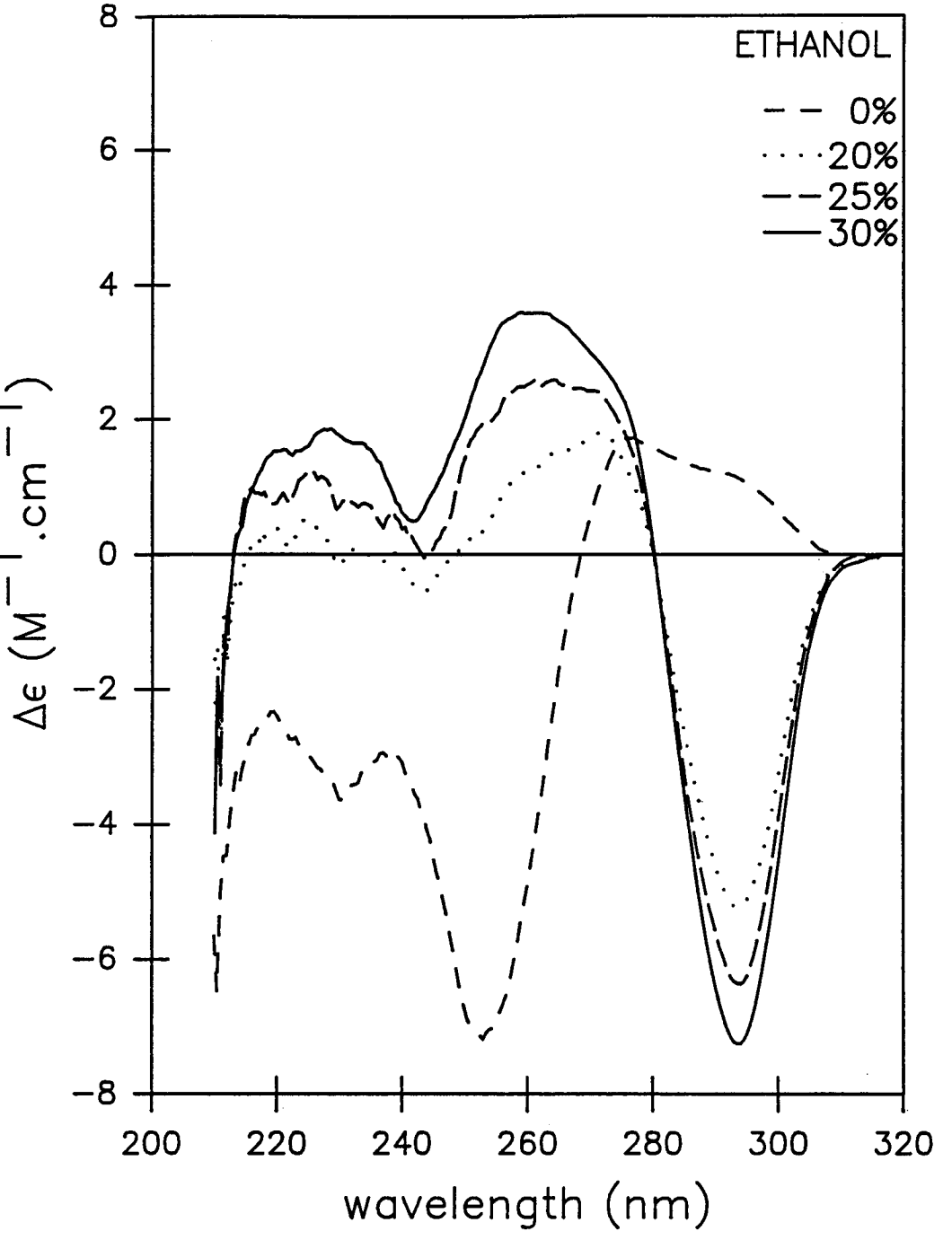
ETHANOL
0%
20%
25%
30%
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.1

coefficients that weight the two significant basis spectra to recreate the original spectra show a smooth titration curve (Figure 3.4). Notice the rapid fluctuations in magnitude and sign of the weighting coefficients for the third basis spectrum, which indicates that this component contributes only noise to the data set and may be neglected.

The short wavelength spectrum of the Z'-form for poly[d(Gm^{5}C).d(Gm^{5}C)] proves that we still have a left-handed form, as it is quite similar to the spectrum for the Z-form (Figure 3.5). It has a negative band around 190 nm with a $\Delta\varepsilon$ of -50. We can see that the intensity of the near UV CD bands changes strikingly during ethanol titration, while the vacuum UV CD bands change little. This indicates that the major change is due to base and sugar-phosphate interactions in the Z'-form rather than base interactions (Sprecher et al., 1979).

It is very interesting to compare the conditions required for the Z-form and Z'-form in poly[d(GC).d(GC)] with the conditions for creating them in poly[d(Gm^{5}C).d(Gm^{5}C)]. For poly[d(GC).d(GC)], 60% ethanol is required for the Z-form, while for poly[d(Gm^{5}C).d(Gm^{5}C)] only 30% ethanol is needed. This indicates that poly[d(Gm^{5}C).d(Gm^{5}C)] favors the Z-form over poly[d(GC).d(GC)]. Both polymers change to the Z'-form at about the same percentage of alcohol, but multivalents are required for the conversion of poly[d(GC).d(GC)] (Razin and Riggs, 1980).

Figure 3.2 Smoothed CD spectra of the Z- to Z'-transition with ethanol for poly[d(Gm^{5}C).d(Gm^{5}C)]. The smoothing methods are described in the Materials and Methods section.

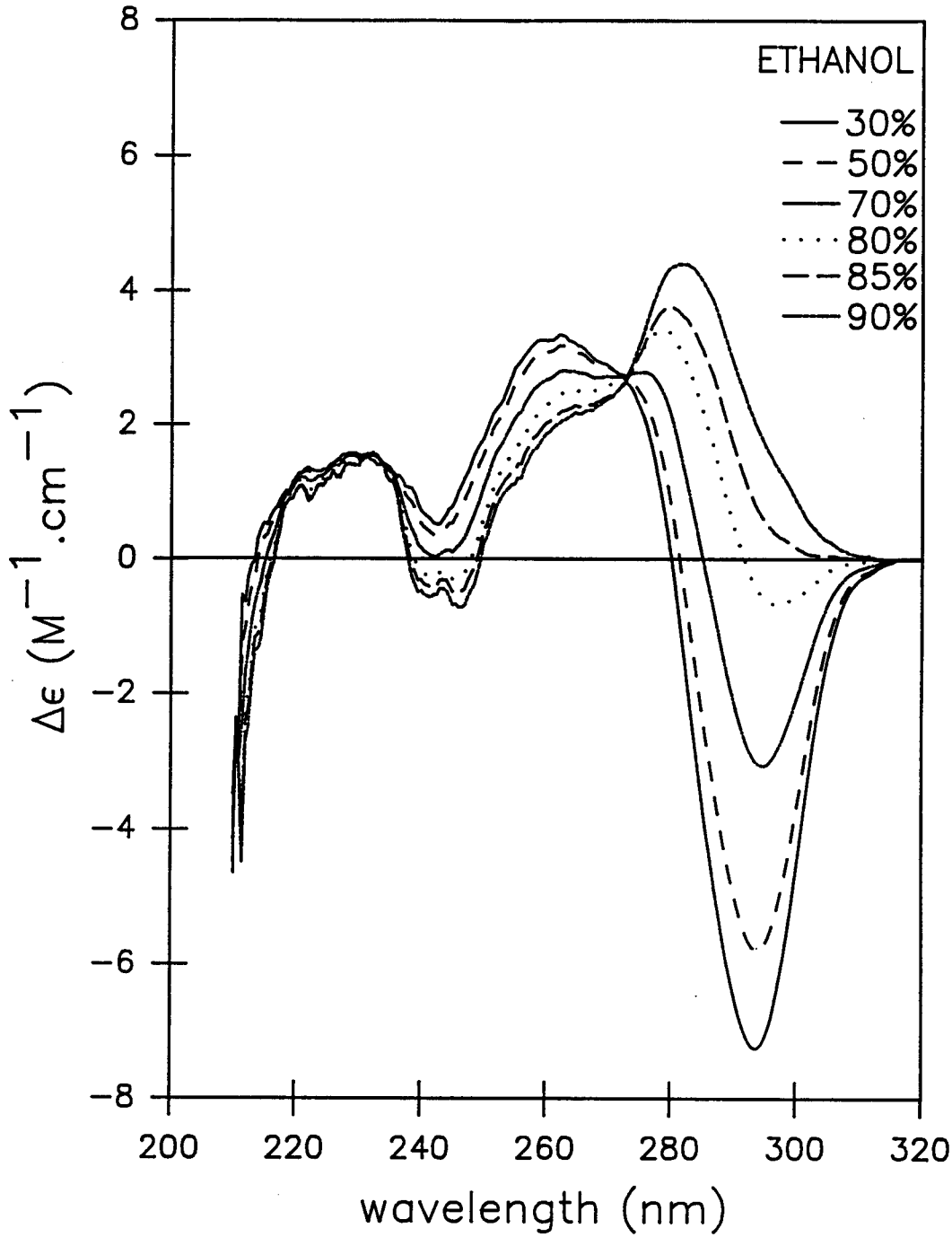
ETHANOL
30%
50%
70%
80%
85%
90%
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
wavelength (nm)
Δε (M⁻¹.cm⁻¹)

Figure 3.2

Figure 3.3 Results of SVD analysis for the ethanol titration of poly[d(Gm^{5}C).d(Gm^{5}C)]: the three most important basis spectra produced by SVD analysis of nine unsmoothed titration spectra. The most important basis vector is the most intense with a negative band at 292 nm. The second basis vector has a positive band around 282 nm. The third most important basis vector is mostly noise and runs along the zero line.

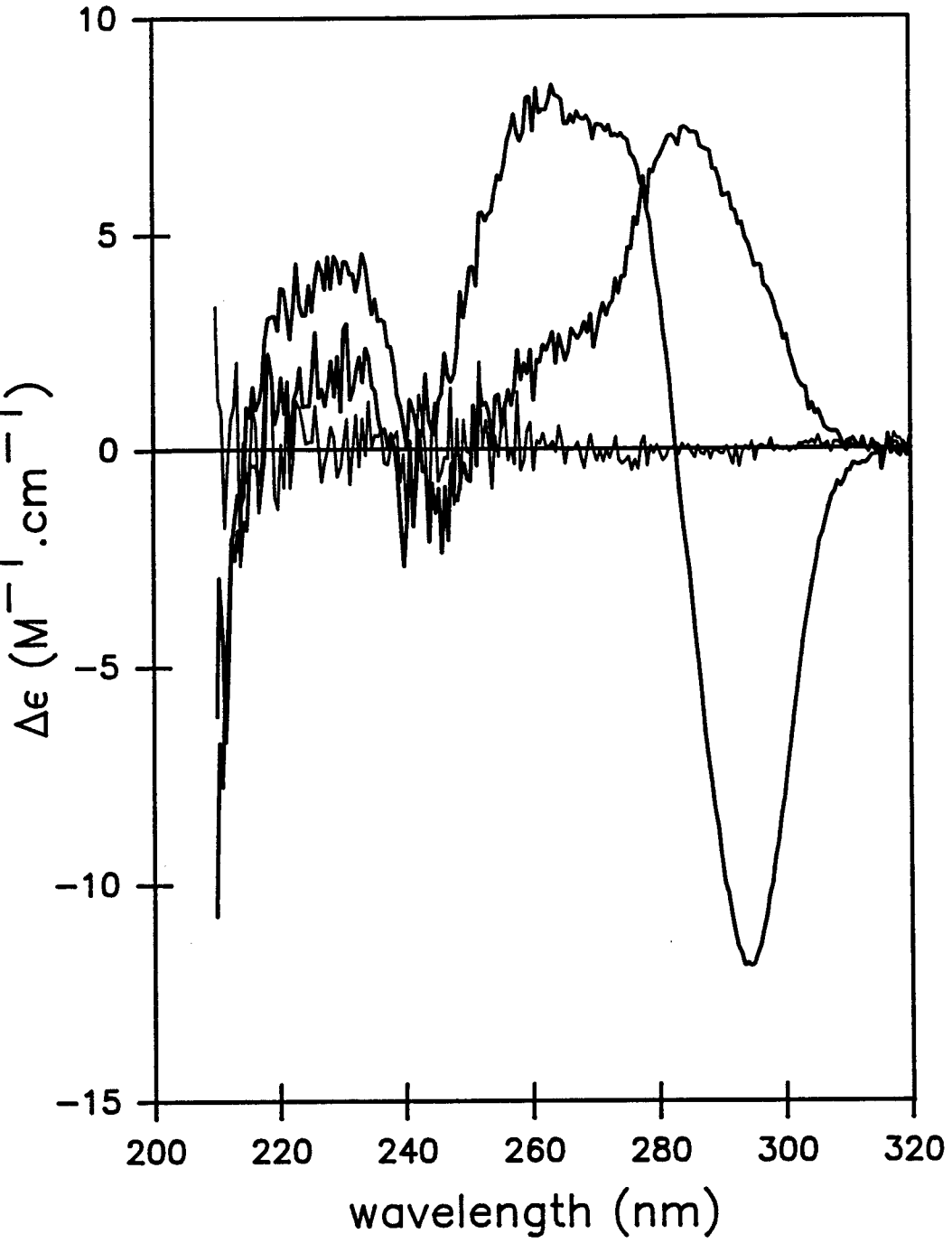

Figure 3.3

Figure 3.4 The three most significant column vectors in matrix **V** corresponding to the basis spectra shown in Figure 3.3. Symbols represent observed values, and smooth curves were drawn by fitting a polynomial regression to the data: (◊) is V_1; (●) is V_2; (▵) is V_3.

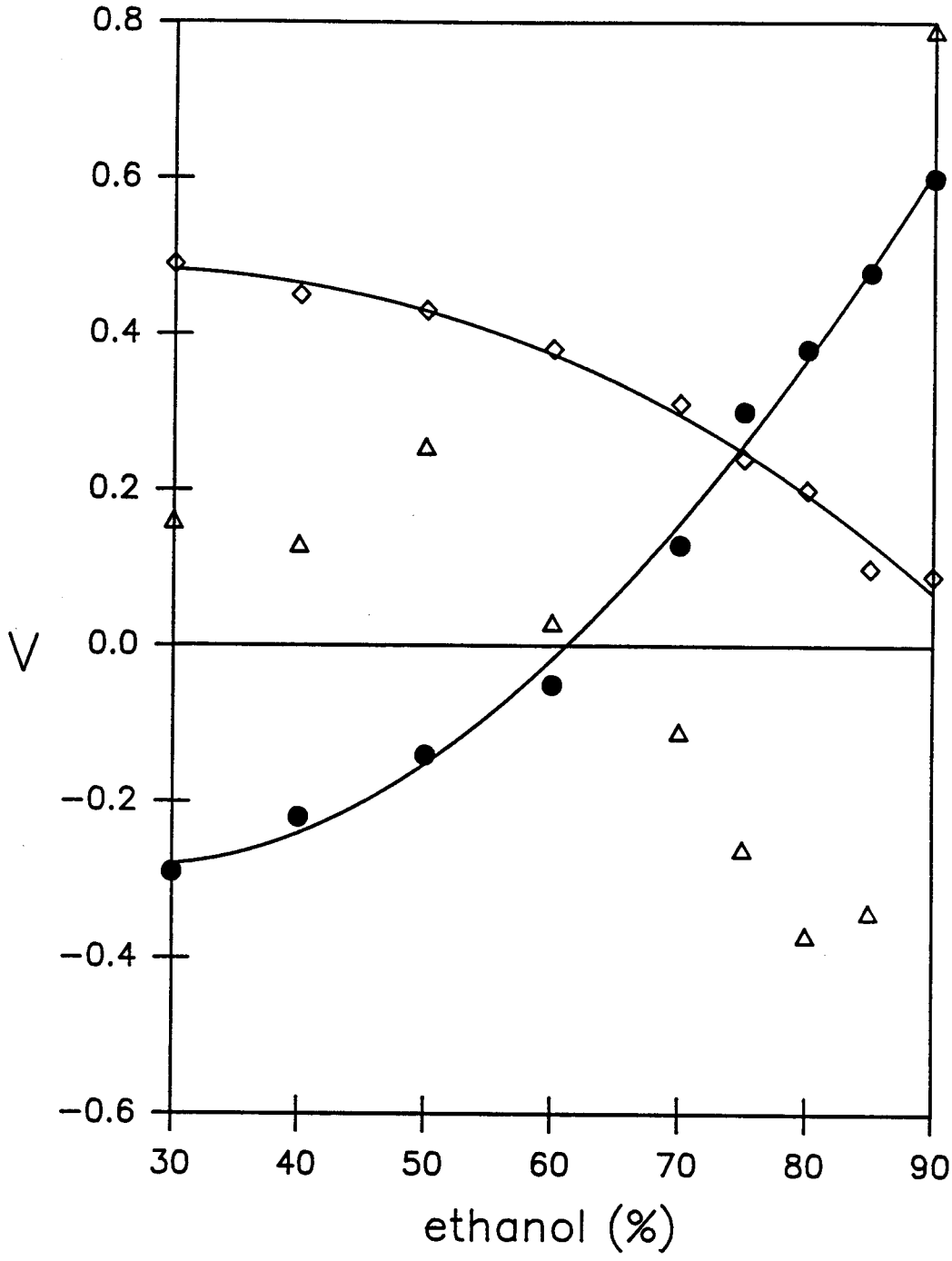

Figure 3.4

Figure 3.5 CD spectra of Z- and Z'-forms extended into the vacuum ultraviolet region.

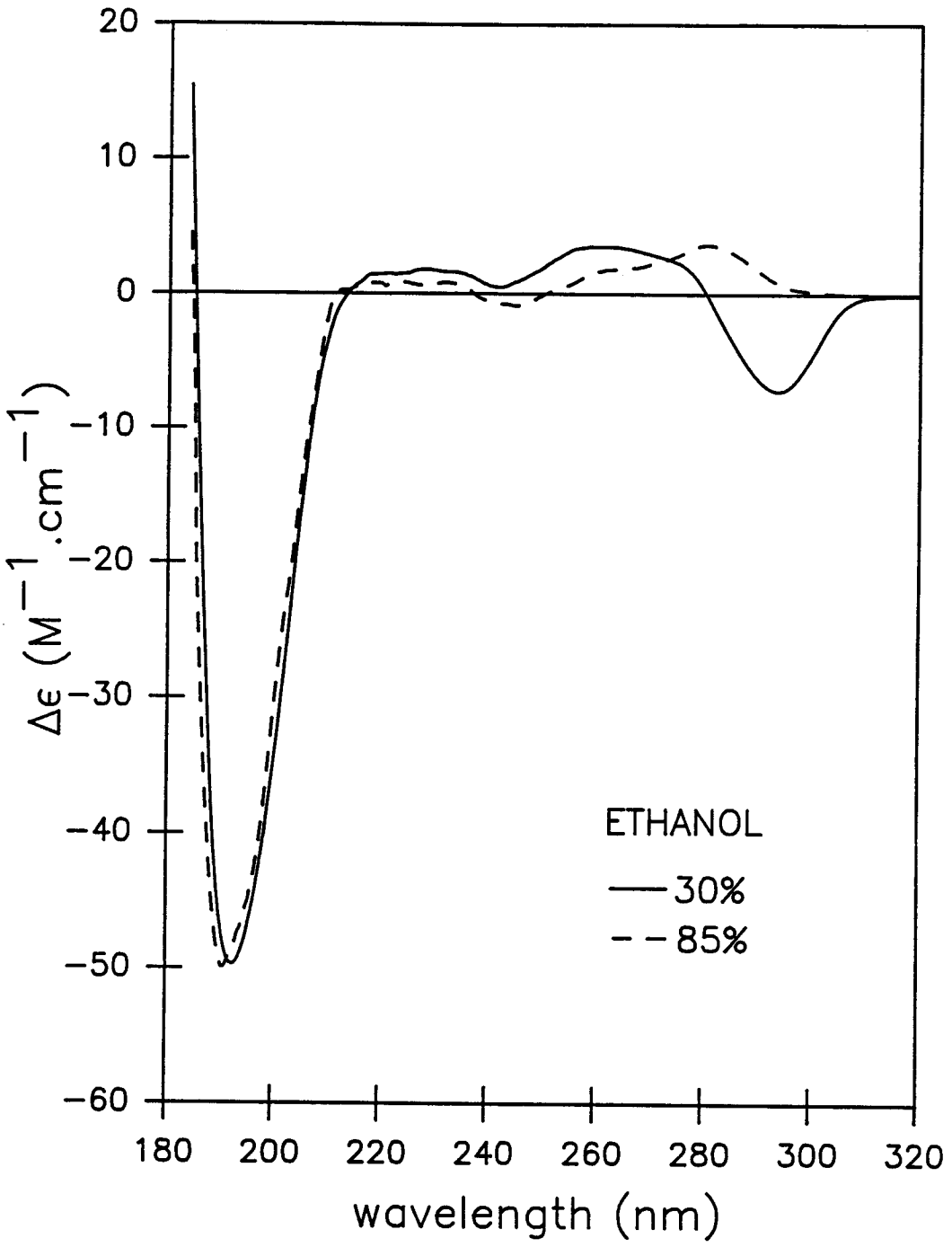
20
10
0
−10
−20
−30
−40
−50
−60
Δε (M⁻¹.cm⁻¹)
180
200
220
240
260
280
300
320
wavelength (nm)
ETHANOL
30%
85%

Figure 3.5

Divalent Ion Titration

Since multivalent ions are necessary to produce the Z'-form of poly[d(GC).d(GC)] in a dehydrating environment (Razin and Riggs, 1980), we studied the effect of metal ions on poly[d(Gm^5C).d(Gm^5C)]. For the Z-form in 30% ethanol we found that when the ratio of Ca^{2+} to nucleotide (r) increases from 0.1 to 4.0, the negative band around 295 nm decreases as the polymer assumes the Z'-form. For the r = 0.1 to 1.0 portion (Figure 3.6a), the spectra show an isosbestic point at 277 nm and the change resembles the ethanol titration (Figure 3.2). However, for r > 1.0 the isosbestic point disappears, and the CD in the region from 255 nm to 280 nm increases (Figure 3.6b) instead of continuing to decrease as in the case of ethanol. When r > 4.0, the DNA starts to precipitate so that a full Z'-form is never reached.

In spite of the lack of an isosbestic point throughout the titration, Figure 3.7 shows that only two SVD components contribute to the CD spectra. The singular values are 133.07, 40.58, and 8.73. Therefore the differences in the Ca^{2+} titration spectra suggest that poly[d(Gm^5C).d(Gm^5C)] has different binding sites for Ca^{2+} during the Z to Z' transition. A plot of the weighting coefficients in Figure 3.8 is in contrast with the ethanol titration (Figure 3.4), in that the weighting coefficient for the second basis spectrum has a complex curve. The coefficients for the first most important basis CD spectrum decrease somewhat at the beginning of the titration and then level out, a particularly simple curve. The coefficients of the third most important basis CD spectrum are scattered over the graph, another indication that this is not significant above the noise. The coefficients for the second most important

Figure 3.6a Smoothed CD spectra for the Ca^{2+} titration of poly[d(Gm^{5}C).d(Gm^{5}C)], with ratio r, calcium ions to nucleotide varying from 0.0 to 1.0.

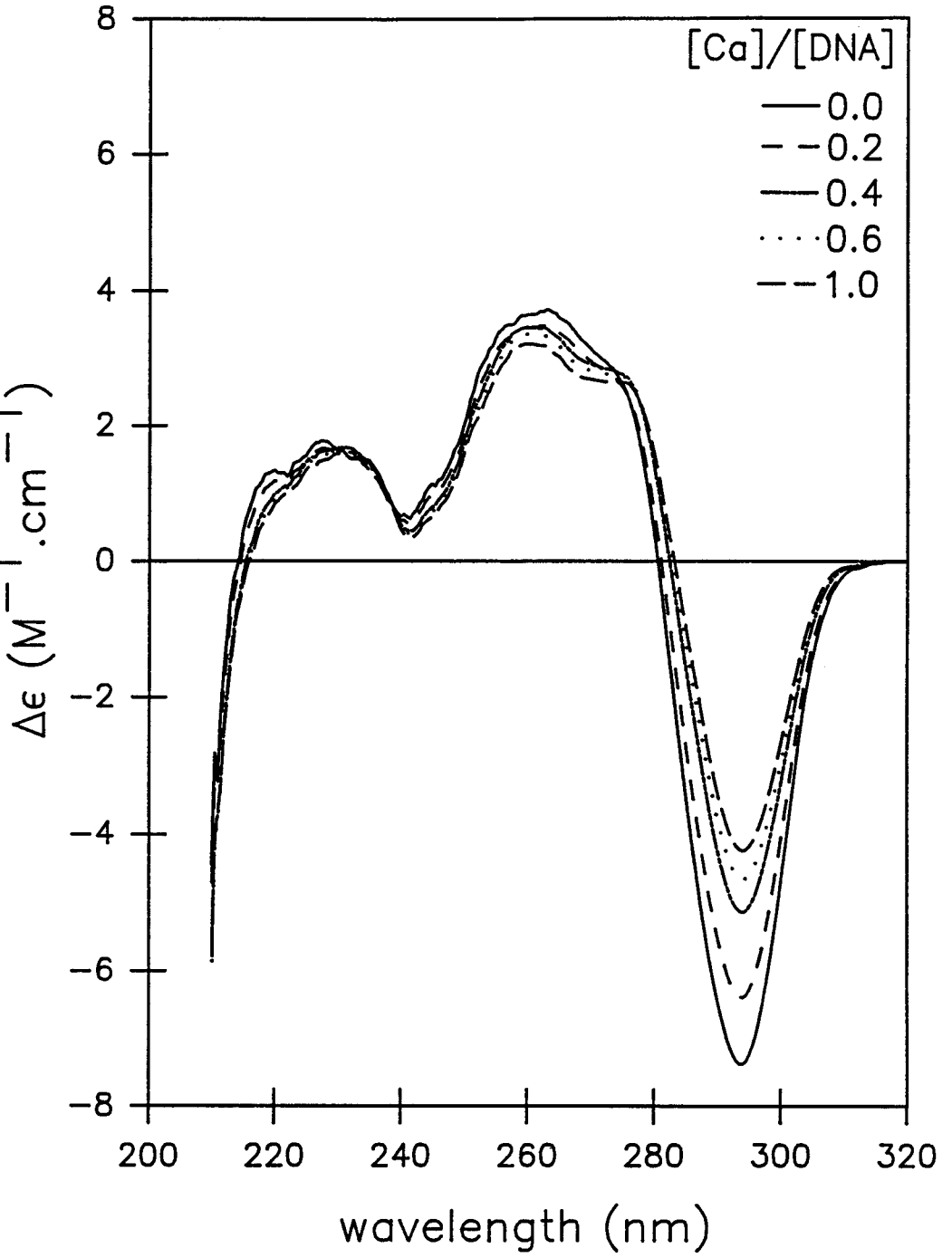
[Ca]/[DNA]
0.0
0.2
0.4
0.6
1.0
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.6a

Figure 3.6b Smoothed CD spectra for the Ca^{2+} titration of poly[d(Gm^{5}C).d(Gm^{5}C)], with r varying from 1.0 to 4.0.

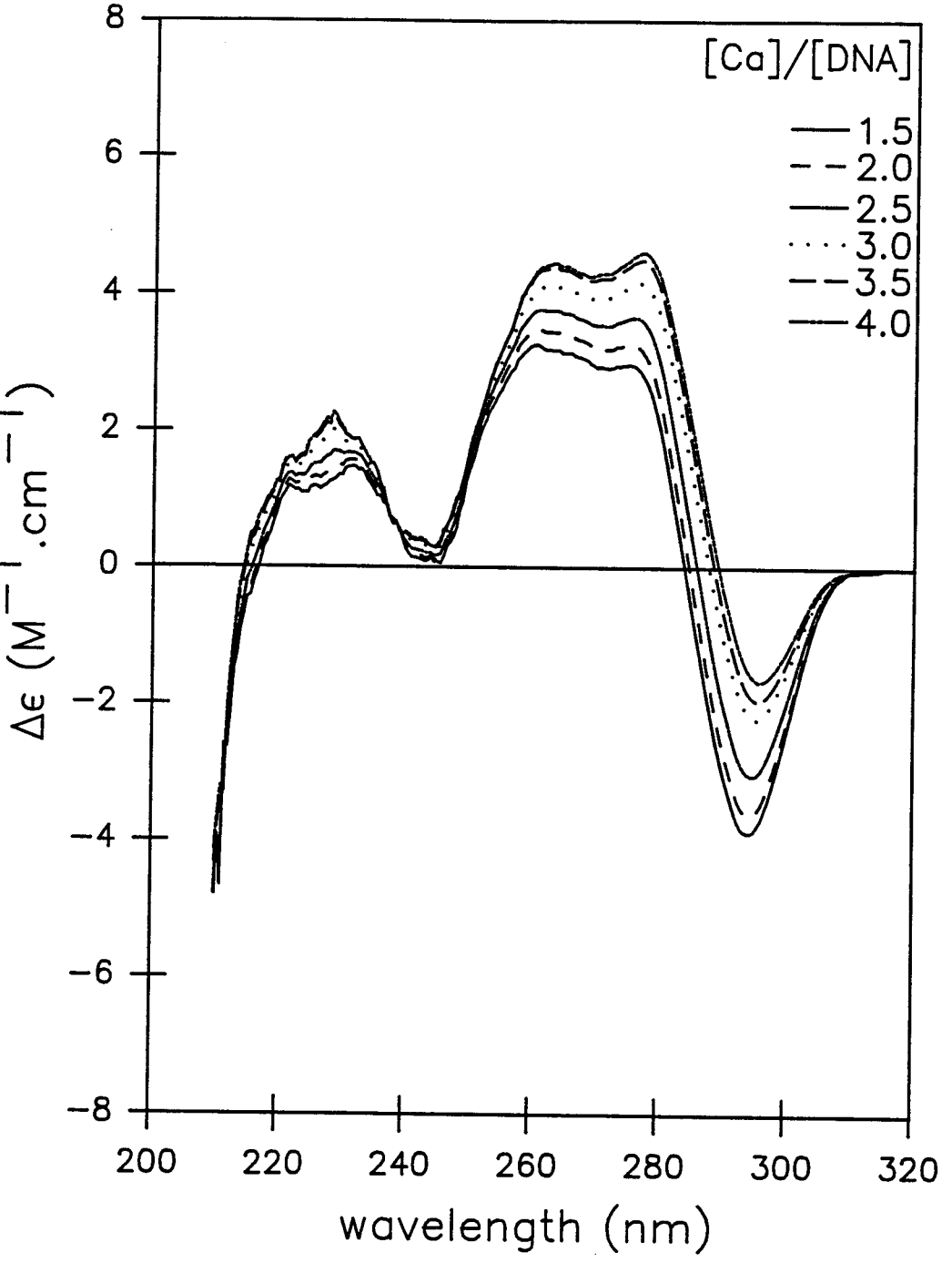
[Ca]/[DNA]
1.5
2.0
2.5
3.0
3.5
4.0
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.6b

basis CD spectrum are a titration curve which represents the variation in CD with added calcium. We can see why there is no isosbestic point in the original data; the curve has two distinct straight line portions, which we interpret as the binding of calcium to two distinct binding sites. It is this type of curve that could be fit with a binding model to determine thermodynamic parameters, such as binding constant, cooperativity, binding stoichiometry, enthalpy, or entropy, as appropriate to the titration. CD spectra for the Ca^{2+} titration of poly[d(Gm^5C).d(Gm^5C)] in 40% EtOH have similar characteristics (data not shown). In contrast, for poly[d(GC).d(GC)], the Ca^{2+} titration spectrum is qualitatively similar to the ethanol titration, and SVD indicates there is only one type of binding site for divalents (Razin and Riggs, 1980).

When we analyze the CD spectra for both the ethanol titration and Ca^{2+} titration of poly[d(Gm^5C).d(Gm^5C)] together, three basis vectors are significant above the noise (Figure 3.9). This indicates that the Z'-forms of poly[d(Gm^5C).d(Gm^5C)] in ethanol and in Ca^{2+} solution are not exactly the same.

Titrations of poly[d(Gm^5C).d(Gm^5C)] were also carried out using the alkaline earth metal magnesium, and the transition metals zinc, cobalt, and nickel. CD spectra for the Mg^{2+} titration resemble the Ca^{2+} titration (Figure 3.10), also an alkaline earth metal. It also does not have a defined isosbestic point, but SVD analysis also shows only two significant basis vectors contributing to the CD spectra. A plot of the weighting coefficients versus r is similar to the calcium titration (Figure 3.8) with the coefficient for the second basis vector indicating two binding sites (not shown).

Figure 3.7 The three most important basis spectra from an SVD analysis of eighteen CD spectra for the Ca^{2+} titration.

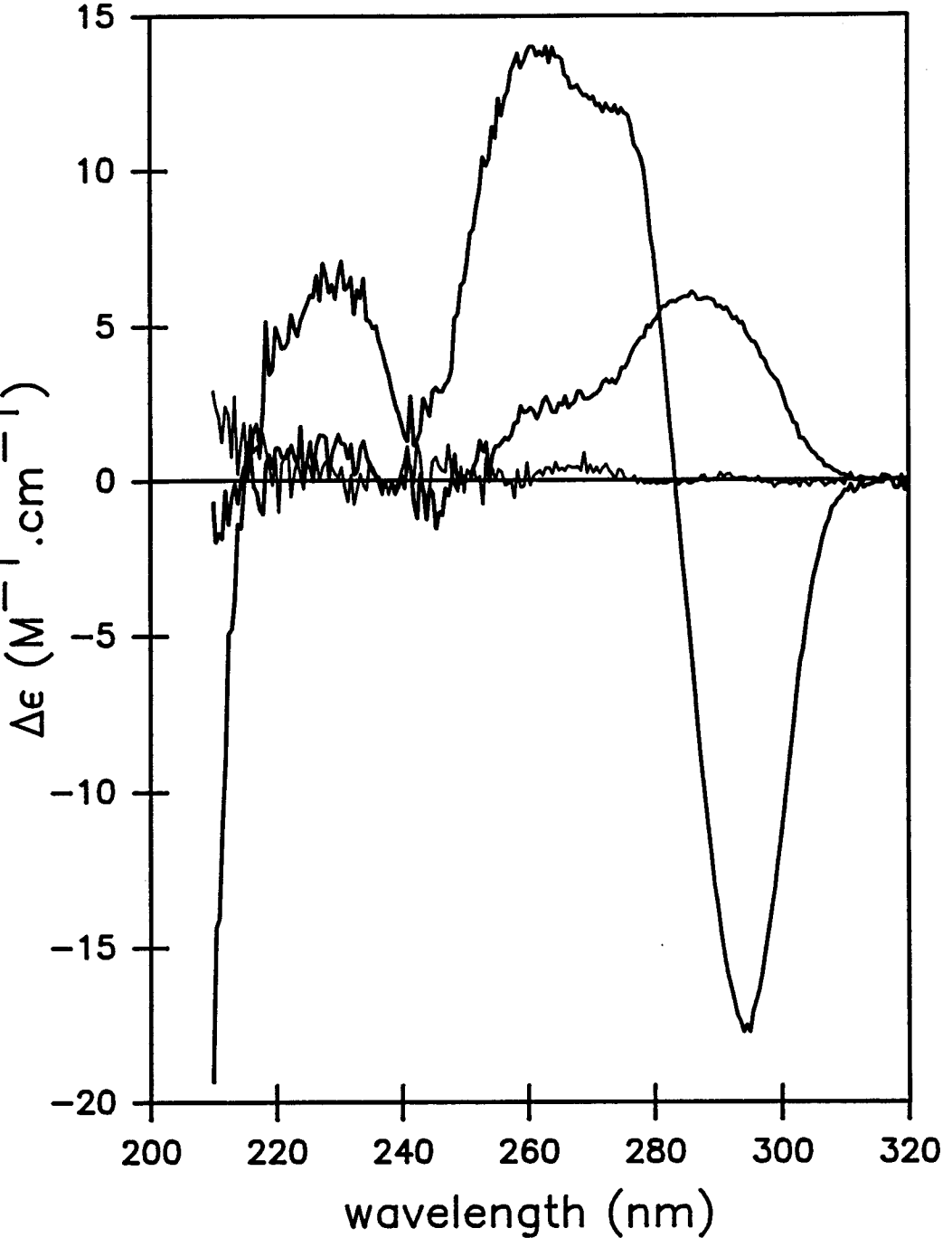
15
10
5
0
−5
−10
−15
−20
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.7

Figure 3.8 The column vectors in **V** corresponding to the three basis spectra shown in Figure 3.7: (◊) is V_1; (●) is V_2; (△) is V_3.

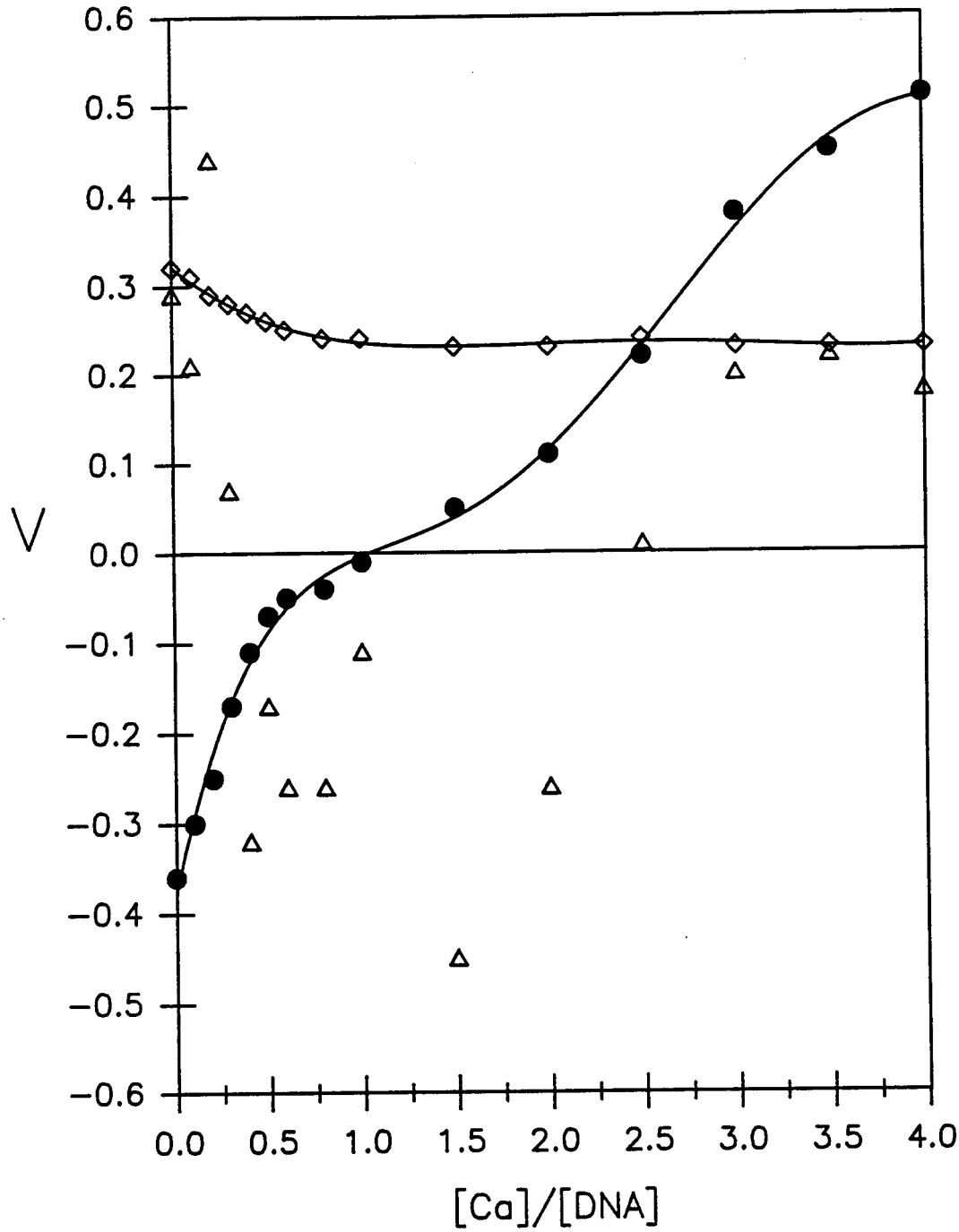
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
V
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
[Ca]/[DNA]

Figure 3.8

Figure 3.9 The three most significant basis spectra from an SVD analysis of CD spectra for the Ca^{2+} and ethanol titrations. Note that in this case all three basis vectors are above the noise.

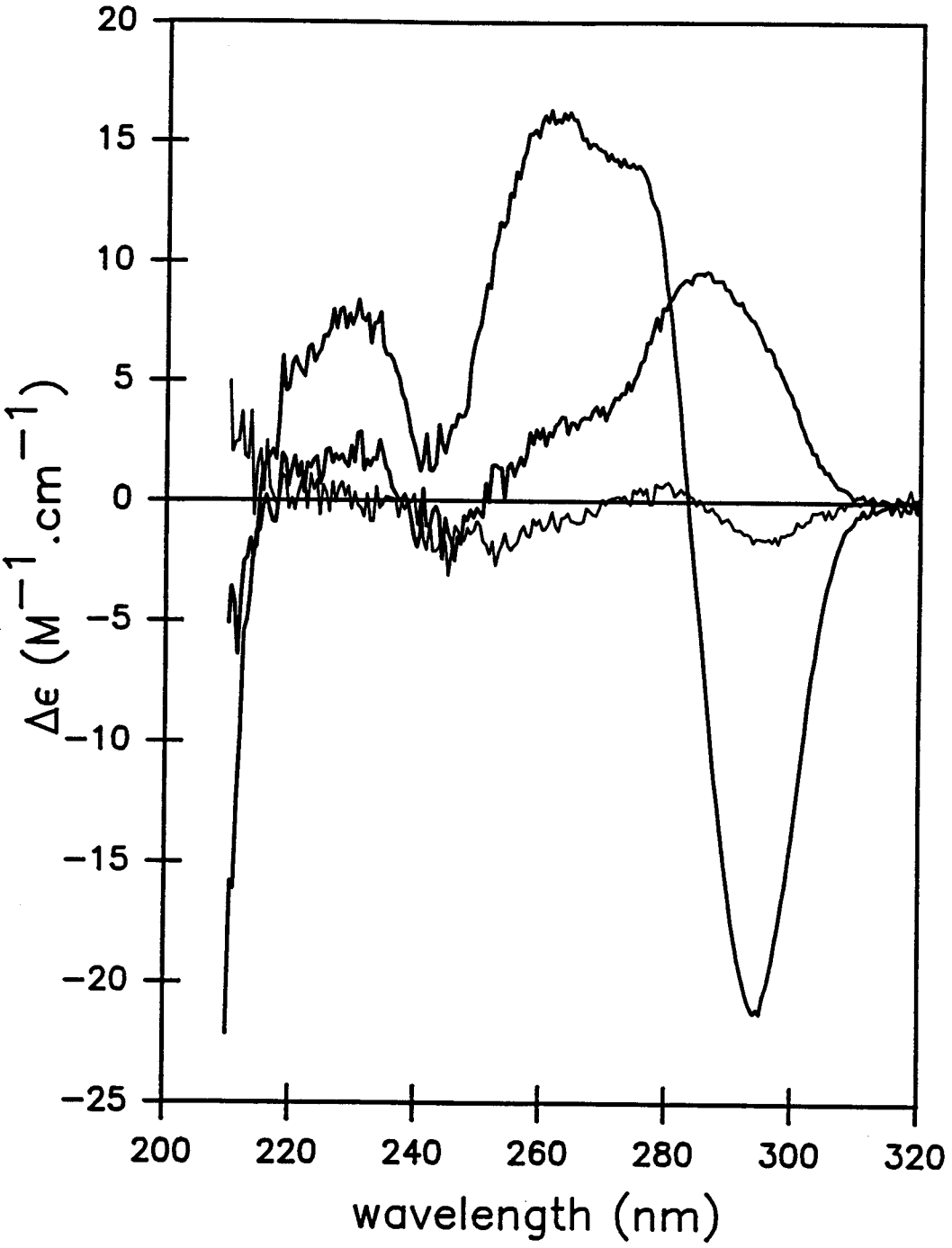
20
15
10
5
0
−5
−10
−15
−20
−25
Δε (M⁻¹.cm⁻¹)
200
220
240
260
280
300
320
wavelength (nm)

Figure 3.9

In contrast to the alkaline earth metals, the transition metals are similar to ethanol in their titration of poly[d(Gm^{5}C).d(Gm^{5}C)] to the Z'-form. Figure 3.11 shows a single isosbestic point at 270 nm and a maximum emerging above 280 nm for the Zn^{2+} titration, although the full Z' spectrum is not reached before the DNA precipitates. This indicates that poly[d(Gm^{5}C).d(Gm^{5}C)] has only one binding site for Zn^{2+}, as does poly[d(GC).d(GC)]. As expected, SVD gives two basis vectors, and the coefficient for the second basis vector is a simple curve (not shown). Co^{2+} and Ni^{2+} titrations have similar CD spectra to the Zn^{2+} titration, however they tend to precipitate the DNA (data not shown).

The results presented here show that related left-handed forms of poly[d(Gm^{5}C).d(Gm^{5}C)] can exist, as is the case for poly[d(GC).d(GC)] (Gessner et al., 1989, Drew et al., 1980, Harder and Johnson, 1990). Increasing the ethanol concentration to 85% or adding various divalent metals produce qualitatively similar changes in the near UV CD spectra of poly[d(GC).d(GC)], in spite of having somewhat different isosbestic points (Harder and Johnson, 1990). This is also true for poly[d(Gm^{5}C).d(Gm^{5}C)] in the case of transition metals, where a single binding site is indicated. However, the alkaline earth metals Ca^{2+} and Mg^{2+} have titration curves that indicate two binding sites to poly[d(Gm^{5}C).d(Gm^{5}C)]. Thus we can see that in divalent ion binding to DNA, the methyl group plays an important role in the structural transition not only from the B- to Z-form (Behe and Felsenfeld, 1981, Fujii et al., 1982), but also from the Z- to Z'-form.

Presumably the conformation found in solution in the absence of

Figure 3.10 Smoothed CD spectra for the Mg^{2+} titration of poly[d(Gm^5C).d(Gm^5C)].

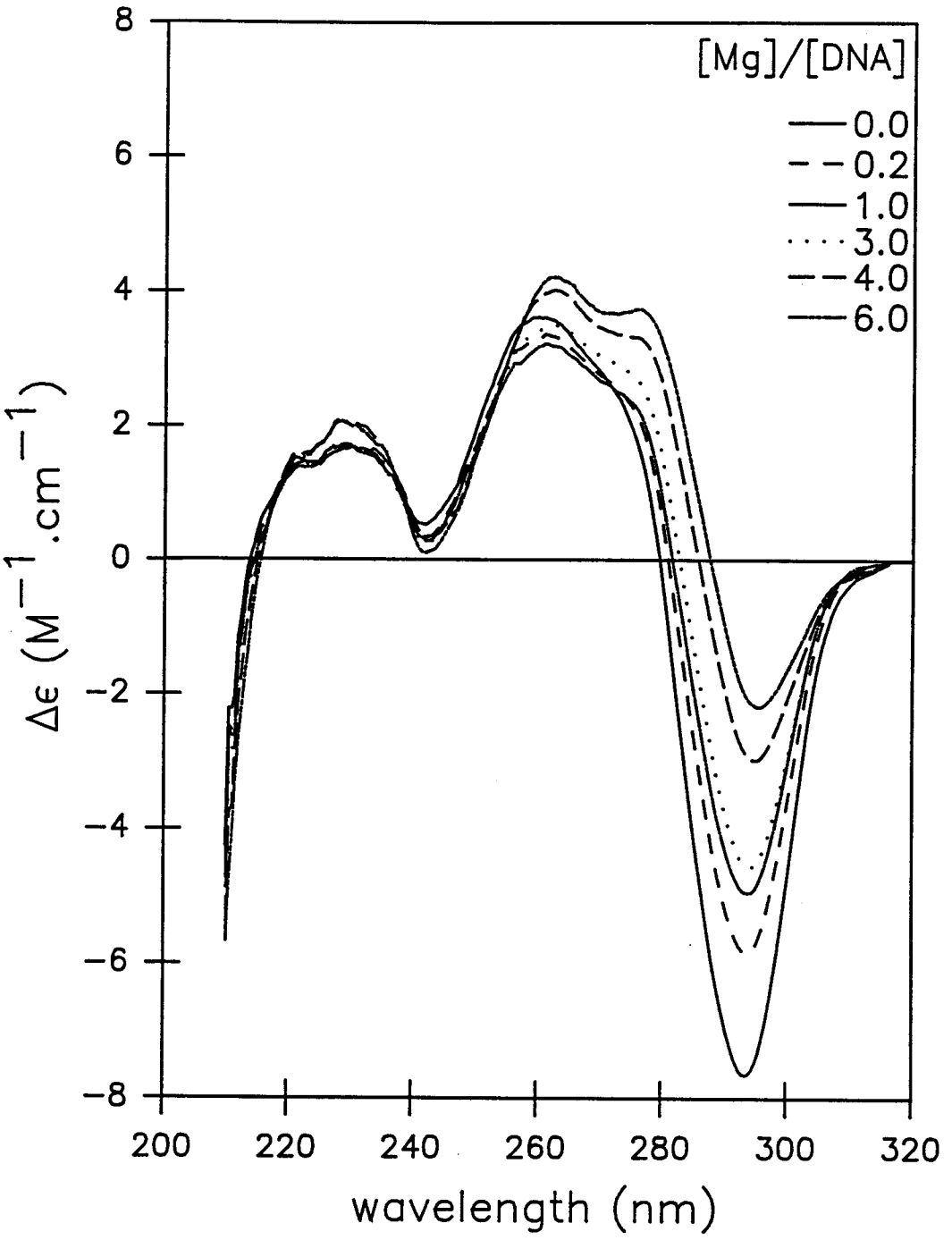
[Mg]/[DNA]
0.0
0.2
1.0
3.0
4.0
6.0
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.10

Figure 3.11 Smoothed CD spectra for the Zn^{2+} titration of poly[d(Gm^{5}C).d(Gm^{5}C)].

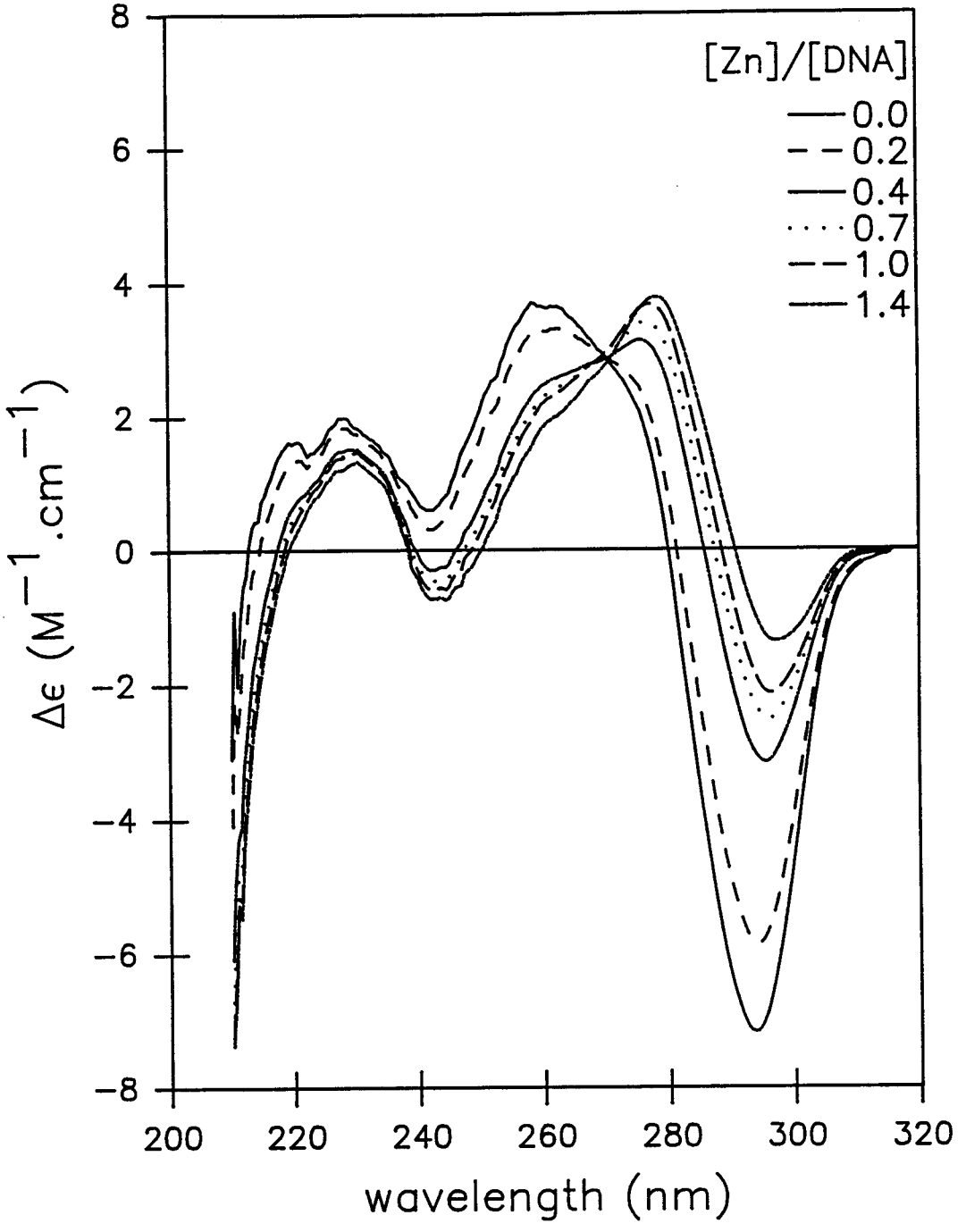
[Zn]/[DNA]
0.0
0.2
0.4
0.7
1.0
1.4
8
6
4
2
0
−2
−4
−6
−8
200
220
240
260
280
300
320
Δε (M⁻¹.cm⁻¹)
wavelength (nm)

Figure 3.11

multivalents, commonly known as the Z-form, is the same conformation as Z in crystals, which has sites with no multivalent ions bound and sites that bind fully hydrated metals. In contrast, sites which have a divalent metal covalently bound to the N of dG (only Mg^{2+} has been studied), and cobalt or ruthenium hexamine hydrogen-bonded directly to the bases, give a different Z-form in crystals, called Z_{II}. Since the Z_{II}-form in crystals requires the binding of multivalent ions, Harder and Johnson (1990) believe it is reasonable to associate the Z_{II}-form with the Z'-form in solution, which also requires the binding of multivalent ions. Since the overall Z forms for crystal $d(GC)_3$ and $d(GmC)_3$, poly[d(GC).d(GC)] and poly[d(Gm^5C).d(Gm^5C)] are similar, it is possible that the Z'-form of poly[d(Gm^5C).d(Gm^5C)] is quite similar to the Z'-form of poly(dGC) with divalent ions. But in the case of the alkaline earth cation metals Ca^{2+} and Mg^{2+}, there are two classes of binding sites differing in their affinity for each of these metals. However, the spectra and therefore the DNA conformations at the two sites are indistinguishable.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. Mark E. Harder for helpful conversations. This work was supported by NSF grant number DMB-8803281 from the Biophysics Program.

REFERENCES

Aubanel, E. E. and Oldham, K. B. (1985) Byte 10, 207-218.

Behe, M. and Felsenfeld, G. F. (1981) Proc. Natl. Acad. Sci. 78, 1619-1623.

Chaires, J. B. (1985) Biochemistry 24, 7479-7485.

Chen, F. (1986) Nucl. Acids Res. 14, 5082-5097.

Chen, G. C. and Yang, J. T. (1977) Anal. Letter 10, 1195-1207.

Devarajan, S. and Shafer, R. H. (1986) Nucl. Acids. Res. 14, 5099-5109.

Doerfler, W. (1983) Annu. Rev. Biochem. 52, 93-124.

Drew, H., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 286, 567-573.

Fujii, S., Wang, A. H., Van der Marel, G., Van Boom, J. H. and Rich, A. (1982) Nucleic Acids Res. 10, 7879-7892.

Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. and Wang, A. H. (1989) J. Biol. Chem. 264, 7921-7935.

Hall, K. B. and Maestre, M. F. (1984) Biopolymers 23, 2127-2139.

Harder, M. E. and Johnson, W. C. Jr. (1990) Nucl. Acids Res., in press.

Hollosi, M., Urge, L., Perczel, A., Kajtar, J., Teplan, I., Otvos, L. Jr. and Fasman, G. D. (1992) J. Mol. Biol. 223, 673-682.

Johnson, W. C. Jr. (1971) Rev. Sci. Instrum. 42, 1283-1286.

Johnson, W. C. Jr. (1985) Physical Optics of Phenomena and Processes in Macromolecular Systems (B. Sedlacek, ed.), pp. 493-506, Gruyter and Co., New York.

Lloyd, D. (1969) Ph.D. Thesis, U. of California, Berkeley.

Noble, B., Daniel, J. W. (1977) "Applied Linear Algebra", Prentice-Hall, Englewood Cliffs, 323-330.

Pohl, F. M. and Jovin, T. M. (1972) J. Mol. Biol. 67, 375-396.

Pohl, F. M. (1976) Nature 260, 365-366.

Razin, A. and Riggs, A. D. (1980) Science 210, 604-610.

Savitzky, A. and Golay, A. E. (1964) Anal. Chem. 36, 1627-1639.

Sprecher, C. A., Baase, W. A. and Johnson, W. C. Jr. (1979) Biopolymers 18, 1009-1019.

Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van der Marel, G., and Rich, A. (1979) Nature 282, 680-686.

Wang, A. H., Quigley, G. J., Kolpak, F. J., Van der Marel, G., Van Boom, J. H. and Rich, A. (1981) Science 211, 171-176.

SECTION IV

BIBLIOGRAPHY

Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961) Proc. Natl. Acad. Sci. USA 47, 1309-1314.

Anson, M.L. and Mirsky, A.E. (1925) J. Gen. physiol. 9, 169.

Arnott, S., Chandrasekara, R., Hukins, D. W. L., Smith, P. J. C., Watts, L. (1974) J. Mol. Biol. 88, 523-33.

Arnott, S., Chandrasekara, R., Biodsall, D. L., Leslie, A. G. W., Ratliff, R. L. (1980) Nature 283, 743-745.

Aubanel, E. E. and Oldham, K. B. (1985) Byte 10, 207-218.

Ausio, J., Zhou, G., and van Holde, K. (1987) Biochemistry 26, 5595-5599.

Baase, W. A. and Johnson, W. C. Jr. (1979) Nucleic Acids Res. 6, 797-814.

Balccrski, J. S., Pysh, E. S., Bonora, G. M. and Toniolo, C. (1976) J. Am. Chem. Soc. 98, 3470-3474.

Barkovsky, E. V. (1982) Acta Biol. Med. Germ. 41, 751-758.

Barrow, C. J. and Zagorski, M. G. (1991) Science 253, 179-182.

Behe, M. and Felsenfeld, G.F. (1981) Proc. Natl. Acad. Sci. 78, 1619-1623.

Behe, M. Zimmerman, S. and Felsenfeld, G.F. (1981) Nature 293, 233.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1965) Nature (London) 206, 757-763.

Blout, E. R. (1962) The dependence of the conformation of polypeptides and proteins upon amino acid composition, in: Ployamino Acids, Polypeptides, and Proteins (M. Stahman, ed.), pp. 275-279, University of Wisconsin Press, Madison.

Bokma, J. T. (1987) Biopolymers 26, 893-909.

Bowie, J. U. , Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) Science 247, 1306-1310.

Bowie, J., Luthy, R. and Eisenberg, D. (1991) Science 253, 164-170.

Brahms, J. and Mommaerts, W. F. H. M. (1964) J. Mol. Biol. 10, 83-88.

Brahms, S., Brahms, J., Spach, G., and Brack, A. (1977) Proc. Natl. Acad. Sci. USA 74, 3208-3212.

Brahms, S., Brahms, J. (1980) J. Mol. Biol. 138, 149-178.

Brandts, J. F. and Hunt, L. (1967) J. Am. Chem. Soc. 89, 4826-4838.

Brennan, R. G., Westhof, E. and Sundaralingam, M. (1986) J. Biomol. Str. and Dyn. 3, 649-665.

Briggs, M. S., Cornell, D. G., Dluhy, R. A., and Gierash, L. M. (1986) Science 233, 206-208.

Bruch, M. D., McKnight, C. J. and Gierasch, L. M. (1989) Biochemistry 28, 8554-8561.

Burgen, A. S. V., Roberts, G. C. K. and Feeney, J. (1975) Nature (London) 253, 753-755.

Burgess, A. W., Ponnuswamy, P. K., and Scheraga, H. A. (1974) Israel J. Chem. 12, 239-286.

Chaires, J. B. (1985) Biochemistry 24, 7479-7485.

Chandrasekaran, R., Wang, M., He, R. -G. Puigjaner, L. C. Byler, M. A., Millaner, R. P., and Arnott, S. (1989) J. Biomol. Struc. Dyn. 6, 1189-1202.

Chen, F. (1986) Nucl. Acids Res. 14, 5082-5097.

Chen, G. C. and Yang, J. T. (1977) Anal. Letter 10, 1195-1207.

Chick, H. and Martin, C. J. (1911) J. Physiol. 43,1.

Chothia, C. (1973) J. Mol. Biol. 75, 295-302.

Chothia, C., Levitt, M., and Richardson, D. (1977) Proc. Natl. Acad. Sci. USA

74, 4120-4134.

Chothia, C., Levitt, M., and Richardson, D. (1981) J. Mol. Biol. 145, 215-250.

Chothia, C. and Janin, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4146-4150.

Chothia, C. and Janin, J. (1982) Biochemistry 21, 3955-3965.

Chothia, C. and Lesk, A. M. (1982) J. Mol. Biol. 160, 309-323.

Chothia, C. (1983) J. Mol. Biol. 163, 107-117.

Chothia, C. (1984) A Rev. Biochem. 53, 537-572.

Chou, P. Y. and Fasman, G. D. (1974a) Biochemistry 13, 211-222.

Chou, P. Y. and Fasman, G. D. (1974b) Biochemistry 13, 222-245.

Chou, P. Y. and Fasman, G. D. (1975) Biochemistry 14, 2536-2541.

Chou, P. Y. and Fasman, G. D. (1978a) Adv. Enzymol. 47, 45-148.

Chou, P. Y. and Fasman, G. D. (1978b) Annu. Rev. Biochem. 47, 251-276.

Cohen, F. E., Richmond, T. J., and Richards, F. M. (1979) J. Mol. Biol. 132, 275-288.

Cohen, F. E., Sternberg, M. J. E., and Taylor, W. R. (1982) J. Mol. Biol. 156, 821-862.

Cohen, F. E., Abarbanel, R. A., Kuntz, I. D., and Fletterick, R. J. (1983) Biochemistry 22, 4894-4904.

Cold Spring Harbor Symp. Quant. Biol. (1983) 47, 1-1229

Contaxis, C. C. and Epand, R. M. (1974) Can J. Biochem., 52, 456-468.

Cook, D. A. (1967) J. Mol. Biol. 29, 167-171.

Crawford, J. L., Kolpak, F. J., Wang, A. H. -J., Quigley, G. H., van Boom, J. H., van der Marel, G., Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 4016-20

Davies, D. R. and Baldwin, R. L. (1963) J. Mol. Biol. 6, 251-255.

Davies, D. R. (1964) J. Mol. Biol. 9, 605-609.

Devarajan, S. and Shafer, R.H. (1986) Nucl. Acids. Res. 14, 5099-5109.

Devynck, M. A., Pernollet, M. C., Meyer, P., Fermandjian, S. and Fromageot, P. (1973) Nature (London), New Biol. 245, 55-58.

Dickerson, R. E. (1990) "What do we really know about B-DNA" in Structure and Methods III (Sarma and Sarma eds.), pp.1-38, Adenine Press.

Dill, K. A. (1990) Biochemistry 29, 7133-7155.

Doerfler, W. (1983) Annu. Rev. Biochem. 52, 93-124.

Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) Nature 286, 567-573.

Drew, H. R. and Dickerson, R. E. (1981) J. Mol. Biol. 152, 723-36.

Dunhill, P. (1968) Biophys. J. 8, 865-875.

Edelman, J. and White, S. H. (1989) J. Mol. Biol. 210, 195-209.

Efimov, A. V. (1979) J. Mol. Biol. 134, 23-40.

Eisenberg D. Wesson M. and Wilcox W. (1989) Predictions of Protein Structure and the Principles of Protein Conformation, (Fasman, G. D. ed.), pp. 635-646, Plenum Press, New York.

Elwell, M. L., Schelleman, J. A. (1977) Biochim. Biophys. Acta 494, 367-383.

Epand, R. M. and Scheraga, H. A. (1968) Biopolymers 6, 1551-1571.

Epand, R. M. (1972) J. Biol. Chem. 247, 2132-2138.

Epand, R. M., Jones, A. J. S., and Schneider, S. (1977) Biochim. Biophys. Acta 491, 296-304.

Fasman, G. D. (1967) Poly-α-amino Acids, Marcel-Dekker, New York, p.505.

Fasman, G. D. (ed.) (1989) Predictions of Protein Structure and the Principles

of Protein Conformation, Plenum Press, New York.

Feigon, J., Wang, A. H.-J., Van der Marel, G. A., Van Boom, J. H. and Rich, A. (1984) Nucleic Acids Res. 12, 1243-1263.

Fujii, S., Wang, A. H., Van der Marel, G., Van Boom, J. H. and Rich, A. (1982) Nucleic Acids Res. 10, 7879-7892.

Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D. (1965) J. Mol. Biol. 12, 60-80.

Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.

Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. and Wang, A. H. (1989) J. Biol. Chem. 264, 7921-7935.

Gessner, R. V., Quigley, G. J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H. and Rich, A. (1985) Biochemistry 24, 237-240.

Gibrat, J.-F., Garnier, J. and Robson, B. (1987) J. Mol. Biol. 198, 425-443.

Gierasch, L. M. (1989) Biochemistry 28, 923-930.

Girod, J. C., Johnson, W. C. Jr., Huntington, S. K., and Maestre, M. F. (1973) Biochemistry 12, 5092-5096.

Goodman, M., Verdini, A. S., Choi, N.S. and Masuda, Y. (1970) in Topics in Stereochemistry (Eliel, E. and Allinger, N.L. ed.), John Wiley, New York, pp. 69-166.

Goodwin, D. C., Brahms, J. (1978) Nucleic Acids Res. 5, 835-50

Graf, L., Csech, G., Barat, E., Ronai, A. Z., Szekely, J. I., Kenesey, A. and Bajusz, S. (1977) Annals M. Y. Acad. Sci. 297, 63-83.

Gratzer, W. B., Beaven, G. H., Rattle, H. W. E., and Bradbury, E. M. (1968) Eur. J. Biochem. 3, 176-283.

Gratzer, W. B. and Beaven, G. H. (1969) J. Biol. Chem. 244, 6675-6679.

Gray, D. M., Taylor, T. N., Lang, D. (1978) Biopolymers 17, 145-57

Guzzo, A. V. (1965) Biophys. J. 5, 809-822.

Hall, K. B. and Maestre, M. F. (1984) Biopolymers 23, 2127-2139.

Harder, M. E. and Johnson, W. C. Jr. (1990) Nucl. Acids Res., in press.

Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V., and Stroud, R. M. (1987) Science 235, 448-455.

Haschemeyer, R. H. and Haschemeyer, A. E. V. (1973) in Proteins, Wiley, New York

Helenius, A. and Simons, K. (1975) Biochim. Biophys. Acta 415, 29-79.

Hennessey, J. P. Jr. and Johnson, W. C. Jr. (1981) Biochemistry 20, 1085-1094.

Hennessey, J. P. Jr., Johnson, W. C. Jr., Bahler, C. and Wood, H.G. (1982) Biochemistry 21, 642-646.

Hennessey, J. P. Jr., Manavalan, P., Johnson, W. C. Jr., Malencik, D. A., Anderson, S. R., Schimerlik, M. I., and Shalitin, Y. (1987) Biopolymers 26, 561-571.

Ho, P. S., Frederick, C. A., Saal, D., Wang, A. H. -J. and Rich, A. (1987) J. Biomol. Str. and Dyn. 4, 521-534.

Ho, S. P. and Degrado, W. F. (1987) J. Am. Chem. Soc. 109, 6751-

Holladay, L. A. and Wilder, P. (1980) Biochim. Biophys. Acta 629, 156-167.

Holley, L. H. and Karplus, M. (1989) Proc. Natl. Acad. Sci. USA 86, 152-156.

Hollosi, M., Urge, L., Perczel, A., Kajtar, J., Teplan, I., Otvos, L. Jr. and Fasman, G. D. (1992) J. Mol. Biol. 223, 673-682.

Ibel, K., May, R. P., Kirschner, K., Szadkowski, H., Mascher, E. and Lundahl, P., (1990) Eur. J. Biochem. 190, 311-318.

Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K., and Poletayev, A. I. (1973) Biopolymers 12, 89-110.

Jirgensons, B. (1977) Biochim. et Biophys 473, 352-358.

Jirgensons, B. (1981) Makromol. Chem. Rapid Commun. 2, 213-217.

Johnson, W. C. Jr. (1971) Rev. Sci. Instrum. 42, 1283-1286.

Johnson, W. C. Jr. (1985) Physical Optics of Phenomena and Processes in Macromolecular Systems (B. Sedlacek, ed.), pp. 493-506, Gruyter and Co., New York.

Johnson, W. C. Jr. (1988) Annu. Rev. Biophys. Biophys. Chem. 17, 145-166.

Kabasch, W. and Sander, C. (1983) FEBS Lett. 155, 179-182.

Kabasch, W. and Sander, C. (1983) Biopolymers 22, 2577-2637.

Kabat, E. A. and Wu, T. T. (1973a) Proc. Natl. Acad. Sci. U.S.A. 70, 1473-1477.

Kabat, E. A. and Wu, T. T. (1973b) Biopolymers 12, 751-774.

Kartha, G., Bello, J., and Harker, D. (1967) Nature (London) 213, 862-865.

Kelly, M. M., Pysh, E. S., Bonora, G. M. and Toniolo, C. (1976) J. Am. Chem. Soc. 99, 3264-3266.

Kendrew, J. C., Dickerson, R. E., Strandbverg, B. E., Hart, R. G., Daview, D. R., Phillips, D. C., and Shore, V. C. (1960) Nature (London) 185, 422-427.

Kennard, O. and Hunter, W. N. (1989) Quart. Rev. Biophys. 22, 327-379.

Killian, J. A., Prasad, K. U., Hains, D., and Urry, P. W. (1988) Biochemistry 27, 4848-4855.

Knecht, R. and Chang, J. Y. (1986) Anal. Chem. 58, 2375-2379.

Kotelchuck, D. and Scheraga, H. A. (1968) Proc. Natl. Acad. Sci. U.S.A. 61, 1163-1170.

Kotelchuck, D. and Scheraga, H. A. (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 14-21.

Kubota, S., Ikeda, K., and Yang, J. T. (1983a) Biopolymers 22, 2219-2236.

Kubota, S., Ikeda, K., and Yang, J. T. (1983b) Biopolymers 22, 2237-2252.

Kullman, W. (1984) J. Med. Chem. 27, 106-115.

Kypr J. and Vorlickova M. (1988) Structure and Expression, 2, DNA and Its Drug Complexes, (Sarma, R. H. and Sarma, M. H. eds.), pp. 105-121, Adenine Press.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., and Hamilton, L. D. (1960) J. Mol. Biol. 2, 38-61.

Lau, S. Y. M., Taneja, A. K. and Hodges R. S. (1984) J. Chromatogr. 317, 129-140.

Lee, C. H., Mizusawa, H., Kakefuda, T. (1981) Proc. Natl. Acad. Sci. USA 78, 2838-42.

Lehrman, S. R., Tuls, J. L. and Lund, M. (1990) Biochem. 29, 5590-5596.

Leslie, A. G. W., Arnott, S., Chandrasekaran, R., and Ratliff, R. L. (1980) J. Mol. Biol. 143, 49-72.

Levin, J. M., Robson, B., and Garnier, J. (1986) FEBS Lett. 205, 303-308.

Levitt, M. and Chothia, C. (1976) Nature 261, 552-558.

Lewis, P. N., Go, N., Go, M., Kotelchuck, D., and Scheraga, H. A. (1970) Proc. Natl. Acad. Sci. U.S.A. 65, 810-815.

Lewis, P. N., Nomany, F. A., and Scheraga, H. A. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 2293-2297.

Lifson, S. and Sander, C. (1980) J. Mol. Biol. 139, 627-639.

Lim, V. I. (1974a) J. Mol. Biol. 88, 857-872.

Lim, V. I. (1974b) J. Mol. Biol. 88, 873-894.

Lloyd, D. (1969) Ph.D. Thesis, U. of California, Berkeley.

Low, B. W., Lovell, F. M., and Rudko, A. D. (1968) Proc. Natl. Acad. Sci. U.S.A. 60, 1519-1526.

Lundahl, P., Greijer, E., Sandberg, M., Cardell, S. and Erikssion, K. -O. (1986) Biochim. Biophys. Acta 873, 20-26.

Maizel, J. V. Jr. (1969) in Fundamental techniques in virology (Habel, K. and Salzman, N. P., eds.), pp. 334-362, Academic Press, New York.

Malencik, D. A., Zhao, Z., and Anderson, S. R. (1990) Anal. Biochem. 184, 353-359.

Manavalan, P. and Johnson, W. C. Jr. (1983) Nature 305, 831-832.

Manavalan, P., Johnson, W. C. Jr., and Modrich, P. (1984) J. Biol. Chem. 259, 11666-11667.

Manavalan, P., Taylor, P., and Johnson, W. C. Jr. (1985) Biochim. Biophys. Acta 829, 365-370.

Manavalan, P., Mittelstaedt, D. M., Schimerlik, M. I., and Johnson, W. C. Jr. (1986) Biochemistry 25, 6650-6655.

Manavalan, P. and Johnson, W.C. Jr. (1987) Anal. Biochem. 167, 76-85.

Maniatis, T., Venable, J. H. Jr., Lerman, L. S. (1974) J. Mol. Biol. 84, 37-64

Manning, M. C. and Woody, R. W. (1987) Biopolymers 26, 1731-1752.

Manning, M. C., Illangeskare, M. and Woody, R. W. (1988) Biophys. Chem. 31, 77-86.

Manning, M. C. (1989) J. of Pharmaceutical and Biomedical Analysis 7, 10, 1103-1119.

Marvin, D. A., Spencer, M., Wilkins, M. H. F., and Hamilton, L. D. (1961) J. Mol. Biol. 3, 547-565.

Matthews, B. W., Siegler, P. B., Henderson, R., and Blow, D. M. (1967) Nature (London) 214, 652-656.

Matthews, B. W. (1975) Biochim. Biophys. Acta 405, 442-451.

Maxfield, F. R. and Scheraga, H. A. (1976) Biochemistry 15, 5138-5153.

Mattice, W. L., Riser, J. M. and Clark, D. S. (1976) Biochemistry 15, 4264-4272.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Green, P., Boyer, H. W.,

Grable, J. and Rosenberg, J. M. (1986) Science 234, 1526-1541.

McLachlan, A. D. (1977) Int. J. Quant. Chem. 13(Suppl. 1), 371-385.

Merutka, G. and Stellwagen, E. (1989) Biochemistry 28, 352-357.

Miller, F. D., Rattner, J. B., van de Sande, J. H. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 571-76

Mitsui, Y., Langridge, R., Shortle, B. E., Cantor, C. R., Grant, R. C., Kodama, M., Wells, R. D. (1970) Nature 228, 1166-1169.

Moser, R., Thomas, R. M. and Gutte, B. (1983) FEBS Lett. 157, 247-251.

Mutter, M. and Hersperger, R. (1990) Angew. Chemie Int. Ed. Eng. 29, 185-187, 1990.

Nagano, K. (1977) J. Mol. Biol. 109, 251-174.

Narayanan, U., Keiderling, T. A., Bonora, G. M. and Toniolo, C. (1986) J. Am. Chem. Soc. 108, 2431-2437.

Nelson, J.W. and Kellenbach, N. R. (1986) Proteins: Struct., Funct., Genet. 1, 211-217.

Nelson, J. W. and Kallenbach, N. R. (1989) Biochemistry 28, 5256-5261.

Nickol, J., Behe, M., Felsenfeld, G. (1982) Proc. Natl. Acad. Sci. USA 79, 1771-75

Nishikawa, K. (1983) Biochim. Biophys. Acta 748, 285-299.

Nishikawa, K. and Ooi, T. (1986) Biochim. Biophys. Acta 871, 45-54.

Nishikawa, K. and Noguchi, T. (1991) Methods in Enzymology 202, 31-44.

Noble, B., Daniel, J. W. (1977) "Applied Linear Algebra", Prentice-Hall, Englewood Cliffs, 323-330.

Osternman, D. G. and Kaiser, E. T. (1985) J. of Cellular Biochem. 29, 57-72.

Palau, J., Argos, P., and Puigdomenech, P. (1982) Int. J. Peptide Protein Res. 19, 394-401.

Patel, D. J., Kozlowski, S. A., Nordheim A., Rich, A. (1981) Proc. Natl. Acad. Sci. USA 79, 1413-17.

Pauling, L., Corey, R. B., and Branson, H. R. (1951) Proc. Natl. Acad. Sci. USA 37, 205-211.

Pauling, L. and Corey, R. B. (1951) Proc. Natl. Acad. Sci. USA 37, 235-285.

Periti, P. F., Quagliarotti, G., and Liquori, A. M. (1967) J. Mol. Biol. 24, 313-322.

Perutz, M. F., Kendrew, J. C., and Watson, M. J. (1965) J. Mol. Biol. 13, 669-678.

Pohl, F. M. and Jovin, T.M. (1972) J. Mol. Biol. 67, 375-396.

Pohl, F. M. (1976) Nature 260, 365-366.

Pongor, S. and Szaley, A. A. (1985) Proc. Natl. Acad. Sci. USA 82, 366-370.

Popoz, E. M. (1980) Mol. Biol. 14, 35-63.

Prevelige, P. E. Jr., Fasman, G. D. (1983) Biochim. Biophys. Acta 739, 85-96

Prothero, J. W. (1966) Biohpys. J. 6, 367-370.

Prothero, J. W. (1968) Biohpys. J. 8, 1236-1255.

Ptitsyn, O. B. (1969) J. Mol. Biol. 42, 501-510.

Ptitsyn, O. B. and Finkelstein, A. V. (1970a) Biofisika 15, 757-768.

Ptitsyn, O. B. and Finkelstein, A. V. (1970b) Dokl. Akad. Nauk. SSSR, 195, 221-224.

Ptitsyn, O. B. and Finkelstein, A. V. (1979) Biofisika 24, 27-30.

Qian, N. and Sejnowski, T. J. (1988) J. Mol. Biol. 160, 865-884.

Rao, P.-F. (1989) Doctoral Thesis, Faculty of Science, Osaka University.

Razin, A. and Riggs, A.D. (1980) Science 210, 604-610.

Reynolds, J. A. and Tanford, C. (1970) Proc. Natl Acad. Sci. USA 66, 1002-1007.

Rich, A., Nordheim, A., and Wang, A. H.-J. (1984) Annu. Rev. Biochem. 53, 791-846.

Riazance, J. H., Baase, W. A. and Johnson, W. C., Jr. (1985) Nucleic Acids Research 13, 13, 4983-4989.

Richardson, J. S. (1981) Adv. Prot. Chem. 34, 167-339.

Richardson, J. S. and Richardson, D. C. (1987) in Protein Engineering (D. L. Oxender and C. F. Fox, eds.), pp.149-163, Liss, New York.

Richardson, J. S. and Richardson, D. C. (1988) Science 240, 1648-1652.

Rhodes, N. J., Mahendrasingam, A., Pigram, W. J., Fuller, W., Brahms, J., Vergne, J., and Warren, R. A. J. (1982) Nature 296, 267-269.

Robertson, T. B. (1918) The Physical Chemistry of the proteins, Longmans, Green and Co., New York.

Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327-356.

Rose, G. D. (1979) J. Mol. Biol. 134, 447-470.

Rose, G. D. and Dworkin, J. E. (1989) Predictions of Protein Structure and the Principles of Protein Conformation, (Fasman, G. D. ed.), pp. 625-634, Plenum Press, New York.

Rosenheck, K. and Doty, P. (1961) Biochemistry 47, 1775-1785.

Saenger, W. 1984. Principles of Nucleic Acid Structure, pp.1-556. New York: Springer

Savitzky, A. and Golay, A.E. (1964) Anal. Chem. 36, 1627-1639.

Schellman, J. A., Lindorfer, M., Hawkes, R., and Grutter, M. (1981) Biopolymers 20, 1989-1999.

Schellman, J. A. (1987) Ann. Rev. Biophys. and Biophys. Biochem. 16,

115-137.

Schiffer, M. and Edmunson, A.B. (1967) Biophys. J. 7, 121-

Schulz. G. E. et al. (1974) Nature 250, 140-142.

Schulz. G. E. and Schirmer, R. H. (1979) Principles of Protein Structures, Springer-verlag, New York

Shirahama, K., Tsujii, K. and Takagi, T. (1974) J. Biochem. (Tokyo) 75, 309-319.

Shneider, A. B. and Edelhoch (1972) J. Biol. Chem. 247, 4992-4995.

Short, K. W., Wallace, B. A., Myers, R. A., Fodor, S. P. A., and Dunker, A. K. (1987) Biochemistry 26, 557-562.

Shortle, D. (1989) J. Biol. Chem. 264, 5315-5318.

Skolnick, J. and Kolinski, A. (1990) Science 250, 1121-1125.

Sprecher, C. A., Baase, W. A. and Johnson, W. C. Jr. (1979) Biopolymers 18, 1009-1019.

Sternberg, M. J. E. and Thornton, J. M. (1978) Nature 271, 15-20.

Sweet, R. M. (1986) Biopolymers 25, 1565-1577.

Takagi, T., Tsujii, K. and Shirahama, K. (1975) J. Biochem. (Tokyo) 77, 939-947.

Taillandier, E. et al. (1985) Advances in Infrared and Raman Spectroscopy Volume 12.

Tamura, Y. and Jirgensons, B. (1980) Arch. Biochem. Biophys. 199, 413-419.

Tanford, C. (1968) Adv. Protein Chem. 23, 121-282.

Tanford, C. (1980) "The Hydrophobic Effect: Formation of Micelles and Biological Membranes" Wiley-Interscience, p.66-68.

Toitskii, G. V. and Zav'yalov, V. P. (1972) J. Mol. Biol. 6, 645-647.

Tunis-Schneider, M. J. B. and Maestre, M. F. (1970) J. Mol. Biol. 52, 521-541.

Unson, C. B., Erickson, B. W., Richardson, D. C. and Richardson, J. S. (1984) Fed. Proc., Fed. Am. Soc. Exp. Biol. 43, 1837.

von Hippel, P. H. and Wong, K. -Y. (1965) J. Biol. Chem. 240, 3909-3923.

von Hippel, P. H. and Schleich, T., The Effects of Neutral Salts on the Structure and Conformational Stability of Macromolecules in Solution., Structure and Stability of Biological Macromolecules, Timasheff, S. N., and Fasman, G. D., Eds. Dekker Inc., N.Y. (1969).

Wallace, B. A., Cascio, M., and Mielke, D. L. (1986) Proc. Natl. Acad. Sci. USA 83, 9423-9427.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., wan Boom, J. H. van der Marel, G., and Rich, A. (1979) Nature 282, 680-686.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H. and Rich, A. (1981) Science 211, 171-76

Wang, J. C. 1969. J. Mol. Biol. 43, 25-39

Watson. J. D. and Crick. F. H. C. (1953) Nature 171, 737-740.

Weber, K. and Osborn, M. (1969) J. Biol. Chem. 244, 4406-4412.

Wells, R. D. (1988) J. Biol. Chem. 263, 1095-1098.

Woody, R. W. (1985) The Peptides 7, 15-114.

Wu, C. C. and Yang, J. T. (1981) Mol. Cell. Biochem. 40, 109-122.

Wu, C. C., Ikeda, K. and Yang, J. T. (1981) Biochemistry 20, 566-570.

Wu, C. C. and Yang, J. T. (1988) Biopolymers 27, 423-430.

Wu, H. (1929) Am. J. Physiol. 90, 562-565.

Wu, H. (1931) Chin. J. Physiol. 5, 321-324.

Wu, H. and Wu, D. (1925) Chin. J. Physiol. 5, 369-374.

Yada, R. Y., Jackman, R. L. and Nakai, S. (1988) Int. J. Pept. Prot. Res., 31, 98-108.

Yang, J. T., Bewley, T. A., Chen, G. C. and Li, C. H. (1977) Proc. Natl. Acad. Sci. USA 74, 3235-3238.

Zehfus, M. H. and Johnson, W. C. Jr. (1981) Biopolymers 20, 1589-1603.

Zehfus, M. H. and Johnson, W. C. Jr. (1984) Biopolymers 23, 1269-1282.

Zimm, B. H. and Bragg, J. K. (1959) J. Chem. Phys. 31, 526-535.

Zimmerman, S.B., Phciffer, B. H. (1980) J. Mol. Biol. 142, 315-30

Zimmerman, S.B. (1982) Ann. Rev. Biochem. 51, 395-427

Zvelebil, M. J., Barton, G. J., Taylor, W. R., and Sternberg, M. J. E. (1987) J. Mol. Biol. 195, 957-961.